
















Temporal profiling of the breast tumour microenvironment reveals collagen XII as a driver of metastasis

Received: 21 May 2021

Accepted: 22 July 2022

Published online: 06 August 2022

 Check for updates

Michael Papanicolaou ^{1,2,3,11}, Amelia L. Parker^{1,3,4,11}, Michelle Yam^{1,3}, Elyse C. Filipe^{1,3,4}, Sunny Z. Wu ^{1,3,4}, Jessica L. Chitty ^{1,3,4}, Kaitlin Wyllie^{1,3}, Emmi Tran^{1,3}, Ellie Mok^{1,3}, Audrey Nadalini^{1,3}, Joanna N. Skhinas^{1,3}, Morghan C. Lucas ¹, David Herrmann ^{1,3,4}, Max Nobis ^{1,3,4}, Brooke A. Pereira^{1,3,4}, Andrew M. K. Law¹, Lesley Castillo¹, Kendelle J. Murphy^{1,3,4}, Anais Zaratzian¹, Jordan F. Hastings¹, David R. Croucher ^{1,4}, Elgene Lim ^{1,4}, Brian G. Oliver^{2,5}, Fatima Valdes Mora ^{6,7}, Benjamin L. Parker⁸, David Gallego-Ortega ^{1,4,9}, Alexander Swarbrick ^{1,3,4}, Sandra O'Toole^{1,4,10}, Paul Timpson ^{1,3,4,12}  & Thomas R. Cox ^{1,3,4,12} 

The tumour stroma, and in particular the extracellular matrix (ECM), is a salient feature of solid tumours that plays a crucial role in shaping their progression. Many desmoplastic tumours including breast cancer involve the significant accumulation of type I collagen. However, recently it has become clear that the precise distribution and organisation of matrix molecules such as collagen I is equally as important in the tumour as their abundance. Cancer-associated fibroblasts (CAFs) coexist within breast cancer tissues and play both pro- and anti-tumourigenic roles through remodelling the ECM. Here, using temporal proteomic profiling of decellularized tumours, we interrogate the evolving matrisome during breast cancer progression. We identify 4 key matrisomal clusters, and pinpoint collagen type XII as a critical component that regulates collagen type I organisation. Through combining our proteomics with single-cell transcriptomics, and genetic manipulation models, we show how CAF-secreted collagen XII alters collagen I organisation to create a pro-invasive microenvironment supporting metastatic dissemination. Finally, we show in patient cohorts that collagen XII may represent an indicator of breast cancer patients at high risk of metastatic relapse.

¹The Garvan Institute of Medical Research and The Kinghorn Cancer Centre, Sydney, NSW, Australia. ²School of Life Sciences, Faculty of Science, University of Technology Sydney, Sydney, NSW, Australia. ³Cancer Ecosystems Program, Garvan Institute of Medical Research, Darlinghurst, NSW, Australia. ⁴School of Clinical Medicine, St Vincent's Healthcare Clinical Campus, Faculty of Medicine and Health, UNSW Sydney, Sydney, NSW, Australia. ⁵Woolcock Institute of Medical Research, Respiratory Cellular and Molecular Biology, The University of Sydney, Sydney, NSW, Australia. ⁶Cancer Epigenetic Biology and Therapeutics, Personalised Medicine, Children's Cancer Institute, Sydney, NSW 2031, Australia. ⁷School of Women's and Children's Health, Faculty of Medicine, UNSW Sydney, Sydney, NSW, Australia. ⁸Metabolic Systems Biology Laboratory, Charles Perkins Centre, School of Life and Environmental Sciences, University of Sydney, Sydney, NSW, Australia. ⁹School of Biomedical Engineering, Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW, Australia. ¹⁰Department of Tissue Pathology and Diagnostic Oncology, Royal Prince Alfred Hospital and NSW Health Pathology, Sydney, NSW, Australia. ¹¹These authors contributed equally: Michael Papanicolaou, Amelia L. Parker. ¹²These authors jointly supervised this work: Paul Timpson, Thomas R. Cox.  e-mail: p.timpson@garvan.org.au; t.cox@garvan.org.au

The progression of all solid tumours is a function of both intrinsic cellular mutational burden and extrinsic effectors provided by the tumour microenvironment¹. Analogous to organ development, breast cancer progression depends on an interplay between the proliferating breast cancer cells and the biochemical and biophysical properties of the extracellular matrix (ECM)². The ECM is a dynamic 3D supramolecular network, essential for coordinating multicellular life. The continually changing ECM during cancer progression shapes and defines the tumour microenvironment, altering both tumour growth and metastatic dissemination.

In cancer, the normal tissue matrix is progressively replaced by the tumour matrix, which is primarily secreted by cancer-associated fibroblasts (CAFs)^{3–7}. Of particular note is collagen I, a major structural component of the ECM that has been extensively studied and its abundance has been associated with both pro-tumourigenic^{8–13}, as well as anti-tumourigenic roles^{14,15} in different solid tumour types. Furthermore, increased collagen I deposition significantly contributes to high mammographic density^{16–19}, a known risk factor for aggressive breast cancer, correlating with poor clinical outcomes^{20–23}. Previous studies on the influence of collagen I abundance in accelerating or suppressing tumour progression point to a more nuanced understanding of the matrix in regulating tumour progression. In particular, the association of collagen I architecture with patient outcome in breast cancer^{12,13,24}, suggests that collagen I architecture and not merely abundance is a critically important regulator of tumour progression.

Collagens are typically secreted and assembled extracellularly during the genesis and repair of the 3D matrix. The deposition and organisation of collagen I fibres and fibrils are regulated by other matrix components, such as through interactions with the proteoglycans decorin²⁵, tenascin X²⁶, the glycoprotein fibronectin²⁷, as well as the family of fibril-associated collagens with interrupted triple helices (FACIT) collagens^{28–30}. Capturing the diversity and complexity of this ECM network requires matrix-specific sample preparation and analytical methods^{31–35}, and has provided fundamental insights into matrix characteristics of the stem cell niche³², tumours³³ and their metastatic sites⁴, for example. Temporal profiling has the potential to reveal dynamic re-organisation of the matrix throughout development and disease progression.

In the context of cancer, the non-selective depletion of the matrix^{36–38}, and/or matrix-producing cells¹⁵, has been shown to have adverse outcomes and can, paradoxically, accelerate tumour progression and metastatic dissemination. An emerging understanding of how the matrix network changes during tumour progression will provide insight into the co-operativity of matrix molecules to regulate cancer cell behaviour. Furthermore, with this in mind, more nuanced approaches focussed on matrix normalisation or re-engineering to restore the composition and architecture of healthy tissues are therefore likely to offer greater promise in the therapeutic setting. Indeed, previous work has shown that collagen cross-linking by lysyl oxidases and changes in tumour stiffness is critically important in driving the progression of breast cancer and that blocking this cross-linking can block tumour progression⁹. Further work has also shown that the biophysical properties of the ECM can also modify treatment response in triple-negative breast cancer³⁹, and agents that modulate stiffness-dependent NF- κ B or JNK activity could enhance therapeutic efficacy in TNBC patients⁴⁰ as well as in other cancer types⁴¹. Therefore, uncovering elements which regulate matrix assembly and organisation, and the role that they play in solid tumour progression will be critical to dissecting the role of the matrix in cancer¹.

However, more work is needed to understand how matrix composition, including the amount of matrix, and in particular the source of secretion (cancer cells versus stromal cells^{3,42}) contribute to the observed pro- and anti-tumourigenic effects on cancer development and progression. It is likely that the balance between pro- and anti-tumourigenic matrix cues likely tips in favour of tumour- and

metastasis-promoting roles as solid tumours progress, and as such understanding the timing of these events will also be critical in designing future interventions.

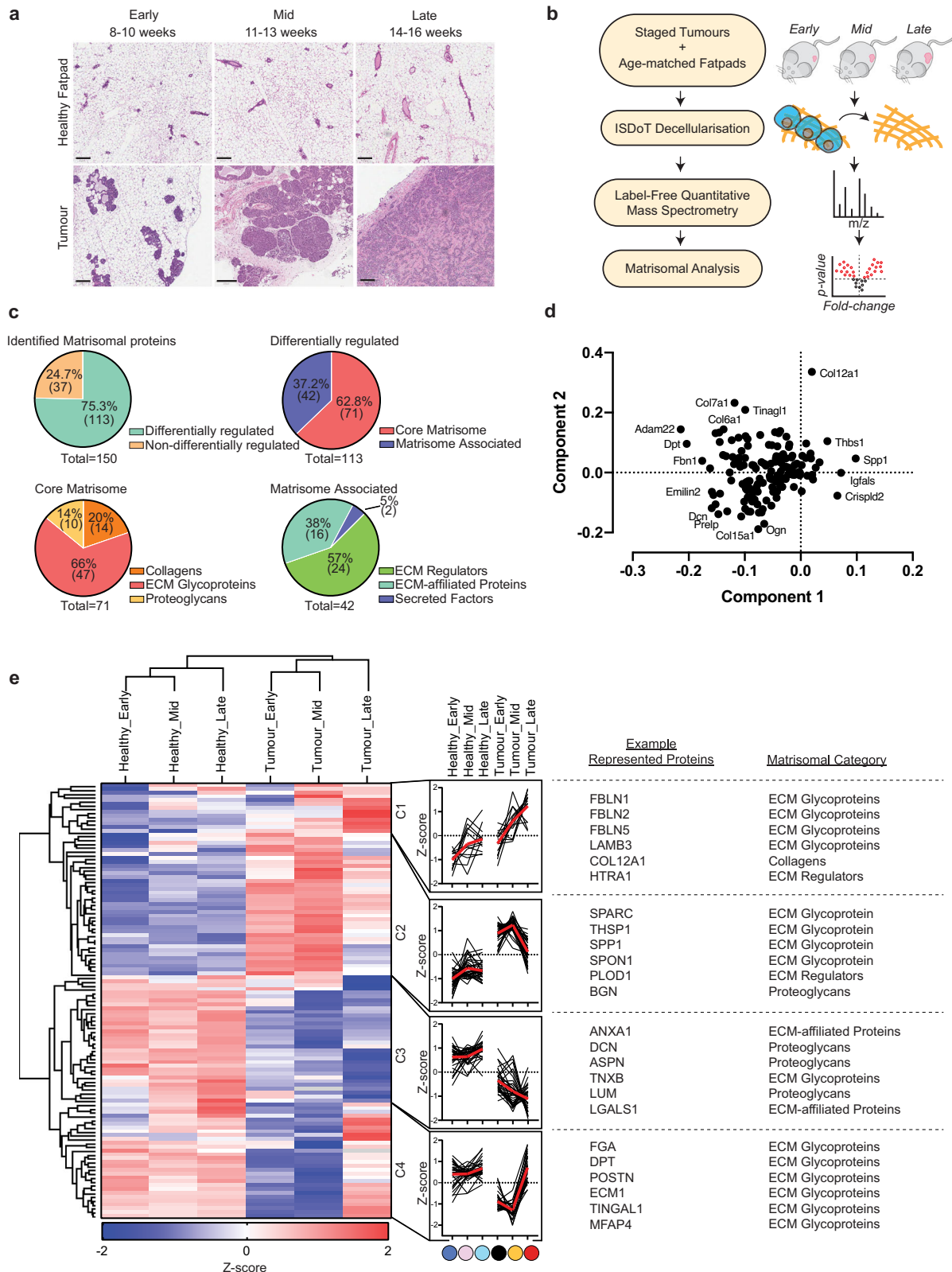
In this work, through coupling of our ISDoT tissue decellularisation technology with quantitative mass spectrometry³¹, we profile the changing tumour matrix during mammary tumourigenesis in the PyMT breast cancer model, identifying temporal changes in the breast cancer matrixome. Through our analysis, we identify collagen XII, a FACIT matrix protein important in regulating collagen I assembly and organisation in load-bearing tissues such as skeletal muscle and tendon as well as the cornea^{30,43–46}, as heavily implicated in mammary tumour progression. Collagen XII has been identified as a marker of poor prognosis in colorectal cancer^{47,48}, although the underlying mechanism is unclear and to date its functional role in breast cancer remains underexplored. We demonstrate that collagen XII is predominantly secreted by CAFs and regulates collagen I fibril organisation to promote cancer cell invasion and breast cancer metastasis. This functional characterisation of collagen XII's importance in breast cancer metastasis is in addition to proteomic data detailing temporal changes in the ECM of breast tumours throughout progression and together warrants further investigation into collagen XII as a potential marker for breast cancer patients at high risk of relapse.

Results

Temporal proteomic profiling of the tumour matrixome

As solid tumours grow, the ongoing remodelling of the ECM leads to the loss of 'normal' tissue matrix and replacement with tumour matrix. This remodelling plays a key role in tumour progression¹. We profiled the temporal changes in ECM of developing breast tumours in the spontaneous polyoma middle-T (PyMT) genetically engineered mouse model (GEMM), from early- (hyperplasia), mid- (adenoma), and late- (metastatic adenocarcinoma) stages, along with age-matched mammary fatpads from healthy mice. Tissues were collected at early (8–10), mid (11–13), or late (14–16 weeks) stages of development (Fig. 1a). As an immunocompetent breast cancer model that mimics the molecular and histological progression of human breast cancer through pre-malignant stages and metastatic dissemination, this model can provide key insight into tumour biology throughout breast cancer progression including a fibroblastic stroma. Gene expression profiling has revealed that the PyMT model recapitulates many aspects of hormone receptor-negative human breast cancers that have the poorest prognosis⁴⁹. Tissues were decellularized in line with our previous work³¹ to enrich ECM proteins before label-free quantitative liquid chromatography-tandem mass spectrometry (LC-MS/MS) (Fig. 1b) was undertaken. A total of 264 matrixome components were consistently detected across all sample datasets (workflow shown in Supplementary Fig. 1A). Of these, 150 were robustly quantified in matched healthy fatpad and tumour samples (Fig. 1c top left, Supplementary Data 1). 113 matrixome elements were determined to be differentially abundant by multi-sample ANOVA (false discovery rate = 0.05), either temporally (early–mid–late), and/or between healthy fatpad and tumour at any given stage of development. Approximately 63% (71/113) were categorised as core matrixome proteins (Fig. 1c top right), comprising proteoglycans, collagens, and glycoproteins (Fig. 1c bottom left). The remaining molecules were matrixome-associated, including secreted factors, ECM-affiliated proteins, and ECM regulators (Fig. 1c bottom right).

Principal component analysis of samples and visualisation of the loadings for Principal components 1 (PC1) and 2 (PC2) identified a number of proteins that most strongly contributed to clustering (Fig. 1d and Supplementary Fig. 1B). Unsupervised hierarchical clustering of \log_2 transformed, median-centred, z-scored data revealed four discrete clusters with distinct temporal profiles (C1–C4) (Fig. 1e left). C1 represents proteins which are progressively upregulated over time in healthy tissues, yet are significantly over-expressed in tumours



and increase further with tumour progression. C2 demonstrates proteins which are consistently upregulated at all stages of tumour development compared to healthy fatpad. C3 includes proteins which are progressively downregulated during disease progression, and C4 represents proteins downregulated in early and mid-stages of development but increase significantly in the later stage compared to normal fatpad. Fisher's exact testing showed no significant association of matrisome categories within each temporal clusters.

Confirming previously published work, we note that the matrix glycoprotein osteopontin (SPP1) is highly abundant at the primary breast tumour site (Fig. 1e right; C2), a matrix element which has been shown to activate mammary fibroblasts into pro-tumourigenic cancer-associated fibroblasts (CAFs)⁵⁰. Furthermore, the proteoglycan decorin (DCN), which is known to be downregulated in breast cancer and has demonstrated both anti-proliferative and anti-metastatic properties^{31,51,52}, was significantly reduced in late-stage tumours

Fig. 1 | Proteomic profiling of decellularized breast tumours reveals dynamic changes in the matrisome. **a** Histological images (H&E) of mammary tumours and age-matched healthy fatpads from early (8–10 weeks), mid (11–13 weeks) and late (14–16 weeks) stages (scale bar = 200 μ m). Representative images of $n = 4$ mice for the tumour group at mid stage and $n = 5$ mice for all other groups. **b** Decell label-free quantitative liquid chromatography-tandem proteomic mass spectrometry (LC-MS/MS) workflow. **c** Breakdown of matrisome classes³⁵ detected in tumours and matched healthy fatpads identified by LC-MS/MS. **d** Principal component analysis (PCA) biplot of tumour and matched healthy fatpad matrisome data. **e** Temporal profiling of mammary tumours and matched healthy fatpad matrisome.

(left) Two-dimensional unsupervised hierarchical clustering of tumours and matched healthy fatpad by Euclidean distance demonstrating 4 clusters (C1–C4) with corresponding profile plot beside each cluster (\log_2 transformed, median-centred z-score of protein abundance values from proteins present in at least 70% of all samples, median of replicate samples from each condition $n = 4–5$). (right) Representative example proteins and corresponding matrisome subcategories. Differentially abundant proteins in LC-MS/MS were determined by multi-sample ANOVA (FDR = 0.05). All data derived from $n = 4$ mice for the tumour group at mid stage and $n = 5$ mice for all other groups. Source data are provided in the Source data file.

compared to healthy fatpad in our data (Fig. 1e right; C3). We also observed the downregulation of tubulointerstitial nephritis antigen-like 1 (TINAGL1) (Fig. 1e right; C4), which has been previously shown to exhibit a tumour suppressive role in breast cancer through binding to epithelial integrin $\alpha 5\beta 1$, $\alpha v\beta 1$, and epidermal growth factor receptor (EGFR), inhibiting focal adhesion kinase (FAK) and EGFR signal transduction⁵³. Finally, we observed downregulation of tenascin X (TNXB) at all stages of tumour progression (Fig. 1e right; C3), which has recently been reported to be downregulated across numerous solid cancer types and associates with poor survival⁵⁴ and may represent a pan-cancer marker. Together these data validate our approach and highlight how key ECM molecules may play important roles in shaping the progression of primary breast tumours, from fostering tumour cell proliferation and invasion into the stroma, migration toward the vasculature, intravasation and vascular dissemination leading to metastatic disease^{1,2}.

Collagen XII shows temporal upregulation during mammary carcinoma development

The treatment of patients with localised primary breast tumours has significantly improved patient outcome in recent years, yet predicting metastatic risk and the treatment of these patients remain a significant clinical problem. When comparing expression changes of matrisome components in healthy fatpad to tumour matrisomes at each stage, we found a number of differentially abundant ECM proteins (Fig. 2a). We were interested in those that were not only robustly detected across all samples, but also those that were lowly and stably expressed in healthy fatpad tissue, and increasingly upregulated as tumours progressed, specifically those in cluster 1 (C1) (Fig. 1e). Collagen XII (COL12A1) demonstrated highly significant up-regulation in tumours at all stages of development compared to healthy controls (Fig. 2a). Furthermore, COL12A1 was identified in our PCA loading analysis as the strongest ECM component contributing to the second principal component (PC2) (Fig. 1d), which separated late- from early-stage tumours (Supplementary Fig. 1B), suggesting it may be strongly associated with tumour progression. Elevated expression of COL12A1 in breast tumours was also supported by our previous ISDoT-based global proteomic profiling of the matrisome in the syngeneic immunocompetent 4T1-Balb/C model where there is an elevated expression of collagen XII in tumours compared to healthy age-matched control³¹. Furthermore, visualisation of the temporal expression profile from early- to mid- to late-stage disease revealed that collagen XII abundance was progressively upregulated as primary tumours developed (Fig. 2b), which was validated by western blot of tissue lysates (Fig. 2c and Supplementary Fig. 1C, D). The spatial expression pattern of collagen XII in primary tumours was evaluated by immunohistochemistry (IHC) and exhibited a predominantly stromal localisation which increased with tumour stage (Fig. 2d, e), confirming its role as a key matrix scaffolding protein^{30,45,46}.

Whilst collagen XII is a known regulator of fibrillar collagen architecture in normal tissues³⁰ and has been correlated with tumour aggressiveness¹², a direct role for this matrix component in regulating breast tumour progression remains to be defined.

Collagen XII expression associates with changing tumour biomechanics

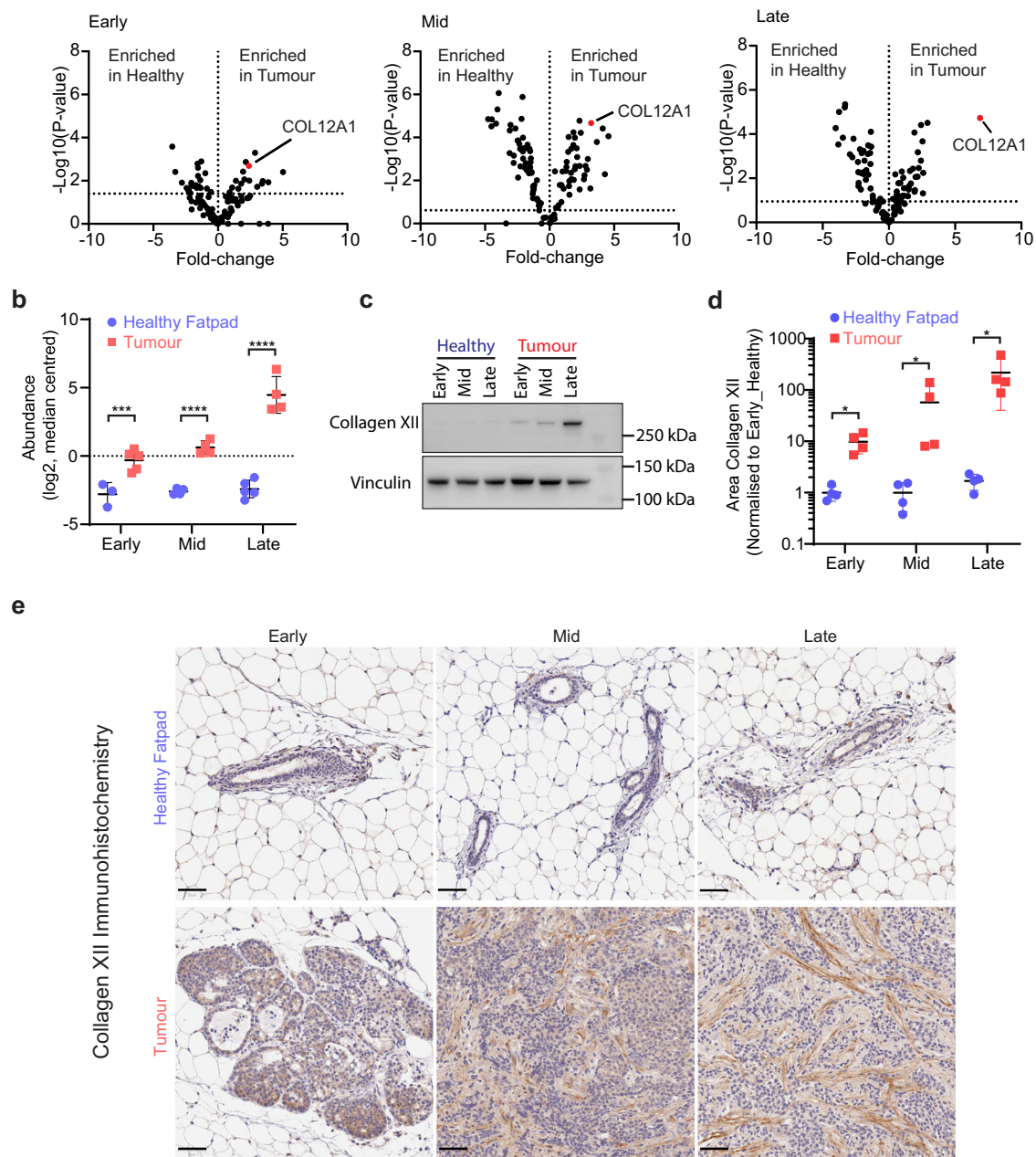
Collagen XII is the largest member of the fibril-associated collagen with interrupted triple helices (FACIT) collagens³⁰. Collagen XII is known to bind to type I collagen fibrils and regulate their organisation and consequently contribute to tissue biomechanics in the cornea, skeletal muscle and tendon tissues^{30,43–46}. Recently, it was elegantly shown that collagen XII expression is associated with high mammographic density in women (a known risk factor of aggressive breast cancer), which is itself associated with increasing matrix stiffness and collagen I bundle thickness²⁰. Collagen XII has also previously been associated with clinical outcomes in gastric cancer⁵⁵. To evaluate the potential role of collagen XII in primary tumour progression, we characterised collagen I/collagen XII spatial organisation over time, as well as tumour biomechanics.

Picrosirius red binds to, and stains fibrillar collagens present within tissues^{56,57}. Using serial picrosirius red stained and collagen XII IHC stained sections from tumours we confirmed that total collagen increased along with collagen XII during tumour progression (Fig. 3a, b and Supplementary Fig. 2A, B). The co-localisation of collagen XII with fibrillar collagen supports the known role of collagen XII in stabilising collagen I fibrils and regulating their 3D organisation in the cornea, skeletal muscle and tendon^{30,43–46}. Using unconfined compression analysis to measure the biomechanical properties of tumours, we confirmed that tumour stiffness (bulk elastic modulus) increased significantly in the late stages of disease (Fig. 3c), correlating directly with the observed increase in collagen XII abundance quantified by mass spectrometry (Fig. 3d).

Exploiting the non-centrosymmetric nature of collagen I fibrils, which allows for label-free imaging and quantification using second-harmonic generation (SHG) multi-photon imaging^{58,59}, we quantified the differences in the organisation of collagen I fibrils and fibres in tumours and healthy fatpad tissues (Fig. 3e). Quantification of SHG peak signal showed an upregulation of collagen I fibre density in the late stage of tumour development (Fig. 3f) matching the increase observed in picrosirius red staining. Quantification of collagen I fibre bundle dimensions revealed an increase in collagen I bundle width as tumours progressed (Fig. 3g, h). Finally, analysis of collagen I fibre orientation³¹ demonstrated an increase in collagen I fibre linearity as tumours progressed from early to mid and late stages (Fig. 3i–l). Together these data confirm a widespread remodelling of fibrillar collagen I organisation during primary tumour progression.

Collagen XII is upregulated in human breast cancer and is associated with a poor prognosis

To determine the clinical relevance of collagen XII in breast cancer patients, collagen XII expression was analysed in RNAseq data from the Cancer Genome Atlas (TCGA) breast cancer cohort (BRCA)⁶⁰. This confirmed significant upregulation of collagen XII across all primary breast tumour types, including TNBC, relative to non-tumour tissue (Fig. 4a). Kaplan–Meier analysis of collagen XII gene expression and patient survival in the same dataset⁶⁰ demonstrated that high collagen XII expression in the primary tumour is significantly associated with both poor overall survival (OS) (Fig. 4b and Supplementary Table 1)

a Healthy Fatpad vs Tumour**Fig. 2 | Collagen XII is upregulated in breast tumours as the disease progresses.**

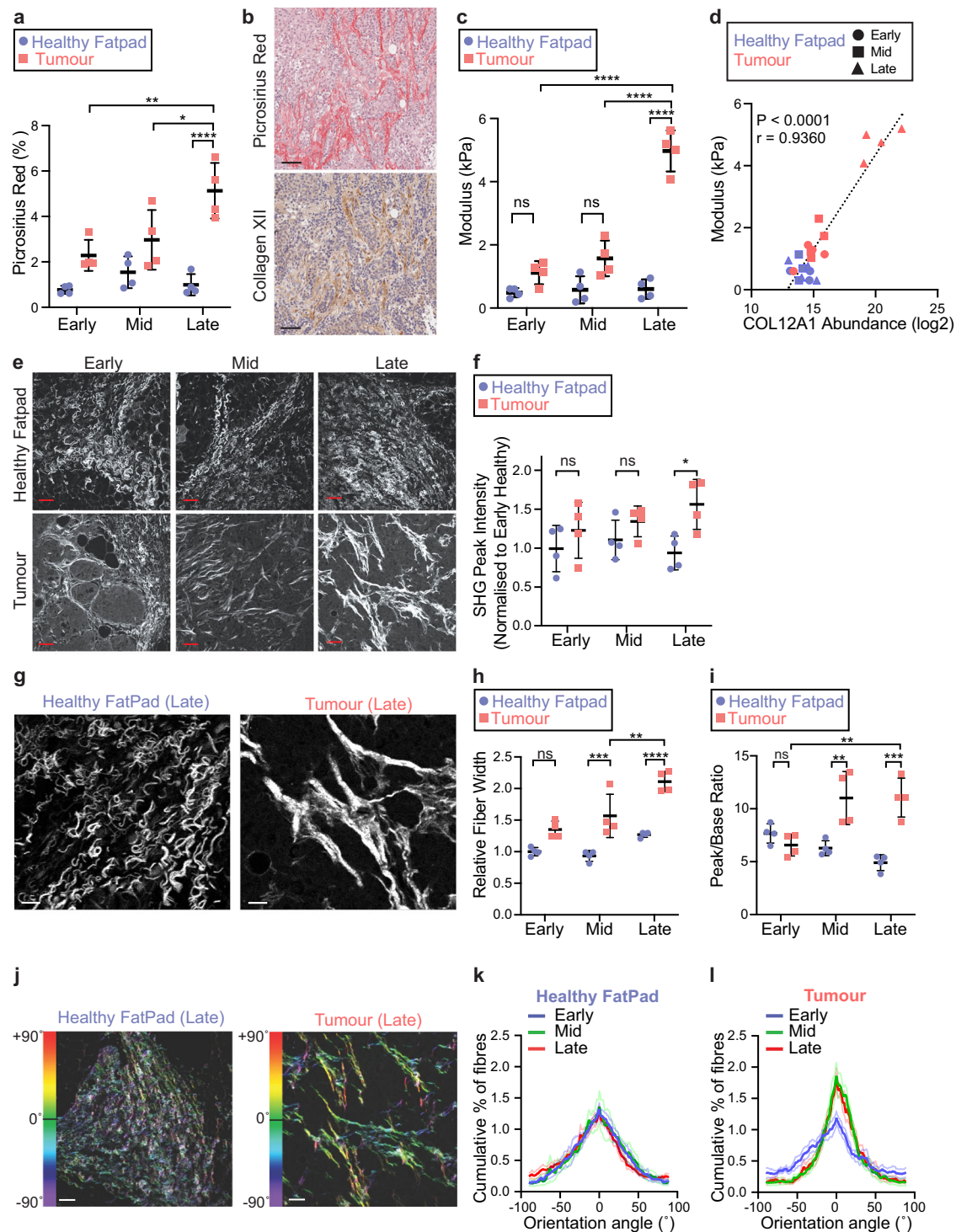
a Volcano plot of differentially abundant matrixome proteins from LC-MS/MS in matched healthy fatpad or mammary tumour at the early, mid or late stages highlighting collagen XII elevation in mammary tumours at each stage. Multi-sample *t* test with FDR correction. Data derived from $n = 4$ mice for the tumour group at mid stage and $n = 5$ mice for all other groups. **b** Collagen XII protein abundance in tumour and matched healthy fatpad tissue, quantified by LC-MS/MS (\log_2 transformed, median centred, collagen XII was detected in $n = 25$ of 30 samples corresponding to tissue samples from $n = 3$ mice in the healthy early stage group, $n = 4$ mice in the healthy mid, tumour mid and tumour late stage groups and $n = 5$ mice from the

healthy late and tumour early stage groups). $***p = 0.0003$, $****p < 0.0001$ One-way ANOVA with Holm–Sidak multiple comparison test. Data are presented as mean \pm SD. **c** Temporal profile of collagen XII protein expression in healthy fatpad and primary tumours via western blot. Representative of $n = 3$ biologically independent samples. **d** Quantification of collagen XII immunohistochemistry (IHC) staining of staged tumours and matched healthy fatpad tissue. Average of $5 \times$ fields of view per tumour in $n = 4$ mice per group. $*p = 0.029$ Mann–Whitney *U*-test (two-sided). Data are presented as mean \pm SD. **e** Representative histological images of collagen XII staining quantified in **d**). Scale bar = $50 \mu\text{m}$ from $n = 4$ biologically independent samples. Source data are provided in the Source data file.

and poor progression-free survival (PFS) (Fig. 4c and Supplementary Table 2). Multivariate analysis of collagen XII expression together with corrections for the clinical covariates age, stage, receptor status and the presence of cancer cell positive lymph nodes revealed that collagen XII expression was the strongest predictor for progression-free survival in early stage (Stages I and II) breast cancer compared to these clinical covariates (Supplementary Table 3). However, collagen XII

expression was not a significant predictor of overall or progression-free survival when tumours across all stages (I–IV) were examined together (Supplementary Tables 3 and 4), suggesting that collagen XII expression is more strongly associated with early-stage progression.

To further confirm these findings, we performed collagen XII immunohistochemistry (IHC) on a patient tissue microarray (TMA) cohort containing 150 breast cancer patients with comprehensive



clinico-demographic and follow-up data⁶¹⁻⁶³. The TMAs contained at least three cores taken from different areas of each patient tumour (Fig. 4d). Our analysis revealed that high collagen XII stromal staining (measured as the % positivity of tumour stroma) was significantly associated with a poor disease-specific-survival (Fig. 4e and Supplementary Table 5), and also with a higher incidence of distant recurrence (Fig. 4f and Supplementary Table 6). Of note is that additional Cox Proportional Hazards multivariate modelling of distant recurrence in this TMA dataset revealed that collagen XII staining is significantly associated with distant recurrence independently of significant clinical risk factors including lymphatic invasion ($p = 0.034$) (Supplementary Table 7). The strong association of collagen XII expression with patient

outcome, and in particular recurrence in early-stage patients, warrants further investigation to determine its potential as a biomarker of aggressive metastatic disease.

Single-cell transcriptomics reveal CAFs underpin collagen XII levels in the tumour microenvironment

The major architects of remodelling of the ECM in solid tumours are known to be CAFs^{64,65}. To assess the source of collagen XII within the primary tumour microenvironment, we analysed single-cell RNAseq (scRNA-seq) of five primary late-stage tumours collected from our *in vivo* models⁶⁶. These data confirmed that the predominant producers of collagen XII (*Col12a1*) are the matrix-secreting subtype of

Fig. 3 | Collagen XII abundance is associated with altered fibrillar collagen architecture and tumour stiffness. **a** Quantification of total picrosirius red stained area in tumour and healthy fatpad tissues. $n = 4$ tissues; $*p = 0.023$, $**p = 0.0022$, $****p < 0.0001$; two-way ANOVA with Tukey's multiple comparison test. Mean \pm SD presented. **b** Representative images of serial histological sections from $n = 4$ mammary tumour tissues stained for fibrillar collagen (picrosirius red), and collagen XII by IHC (scale bar = $50 \mu\text{m}$). **c** Unconfined compression analysis of $n = 4$ tumours and healthy fatpads (mean \pm SD, $n = 4$). $****p < 0.0001$, ns = not statistically significant; early healthy vs. early tumour $p = 0.37$; mid healthy vs. mid tumour $p = 0.051$; two-way ANOVA with Tukey's multiple comparison test. **d** Correlation between collagen XII protein abundance (LC-MS/MS) and bulk modulus of primary tissues (Pearson correlation test, $r = 0.936$, $p < 0.0001$; $n = 24$ tissues). **e** Representative single-plane images of second harmonic generation (SHG) signal acquired from $n = 4$ tumours and healthy fatpad tissues at early, mid and late stages (scale bar = $40 \mu\text{m}$). **f** Quantification of peak SHG multi-photon signal. $n = 4$ tissues

per group; $*p = 0.017$, ns = not statistically significant: early healthy vs. early tumour $p = 0.58$; mid healthy vs. mid tumour $p = 0.57$. Two-way ANOVA with Tukey's multiple comparison test. **g** Representative single-plane images of SHG multi-photon images acquired from $n = 4$ late-stage tumours and healthy fatpad tissues (scale bar = $15 \mu\text{m}$). **h** Quantification of collagen I fibre bundle width in images from **g**. Mean \pm SD of $n = 4$ tissues per group; $**p = 0.0032$, $***p = 0.0006$, $****p < 0.0001$, ns: $p = 0.083$; two-way ANOVA with Tukey's multiple comparison test. **i** Quantification of collagen I fibril orientation peak to baseline ratio. Mean \pm SD, $n = 4$ tissues per group; $**p < 0.01$, $***p < 0.001$, ns: $p = 0.89$; mid healthy vs. mid tumour $p = 0.0025$, late healthy vs. late tumour $p = 0.0001$, early tumour vs. late tumour $p = 0.0042$; two-way ANOVA with Tukey's multiple comparison test. **j** Representative images of fibril-orientation analysis for collagen I fibres in $n = 4$ tissues (scale bar = $40 \mu\text{m}$). Collagen I fibril orientation distribution for **k** staged healthy fatpad and **l** tumour (\pm SEM, $n = 4$) corresponding to analysis in **i**. Source data are provided in the Source data file.

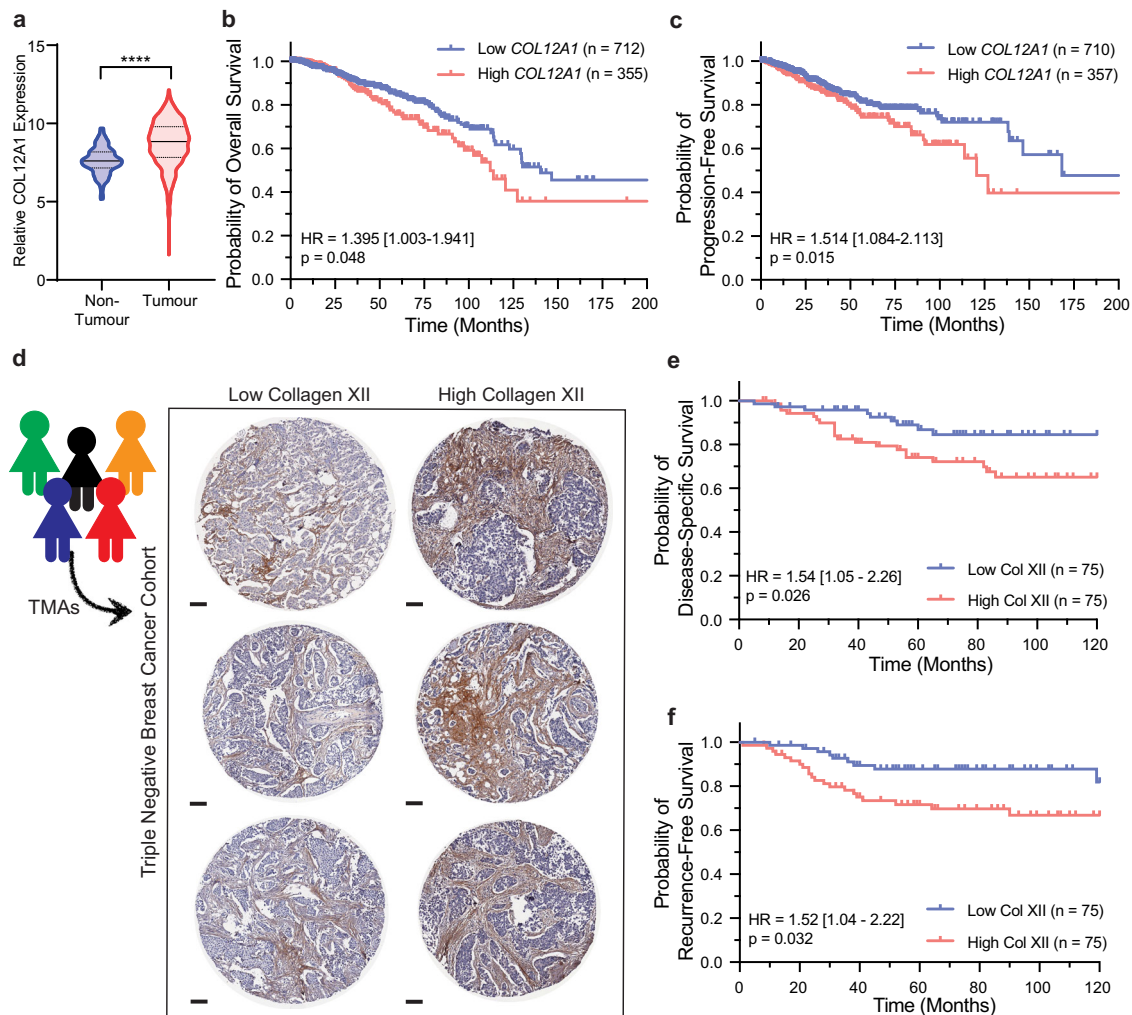
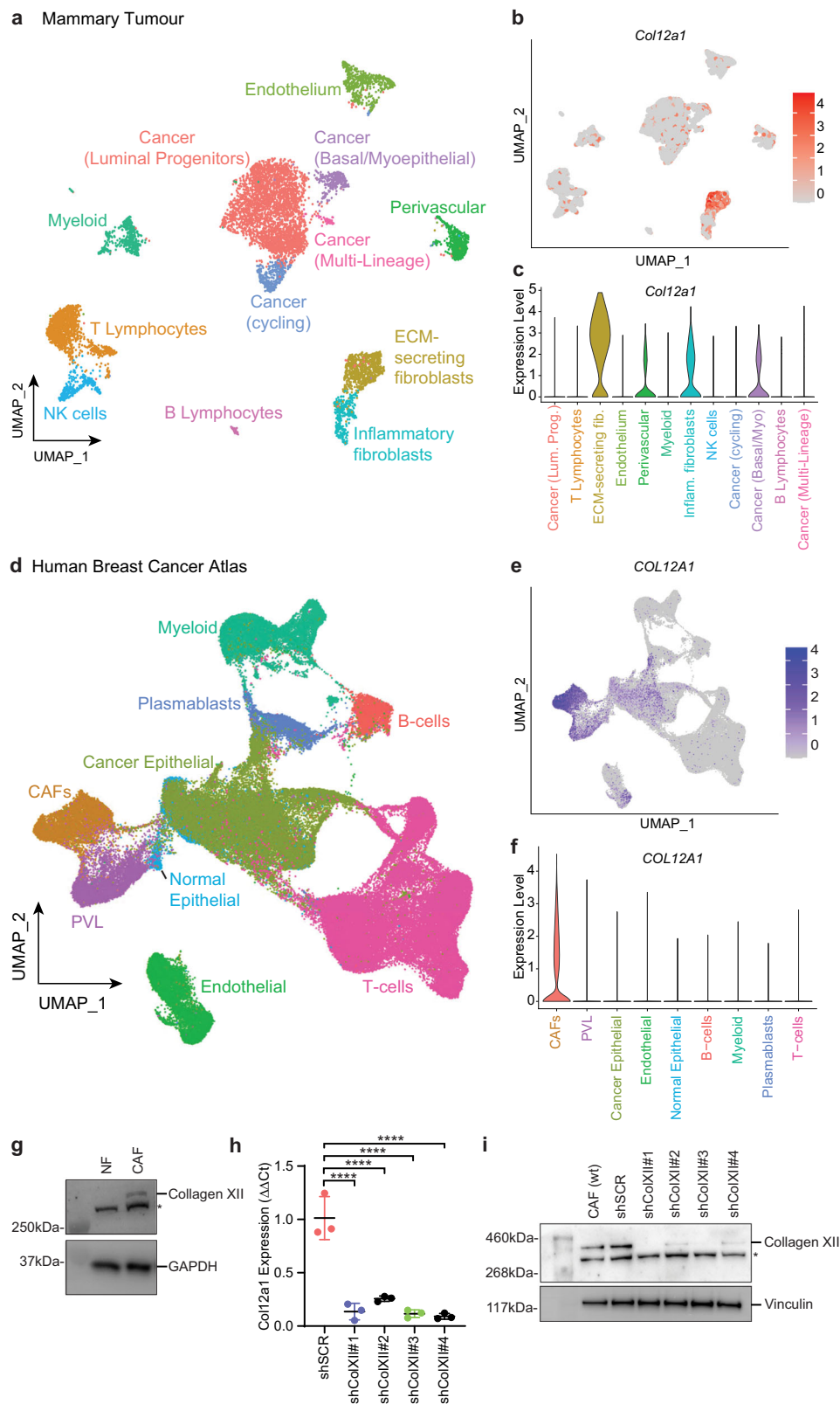


Fig. 4 | Collagen XII expression is associated with poor prognosis in human breast cancer. **a** COL12A1 expression in human non-tumour mammary tissue ($n = 98$) or primary tumour ($n = 1080$) from 'The Cancer Genome Atlas' (TCGA) breast adenocarcinoma cohort (BRCA)⁶⁰ (Median and interquartile range, p value determined by two-sided Mann-Whitney test ($****p < 0.0001$ Mann-Whitney U -test)). **b** Kaplan-Meier overall survival and **c** progression-free survival analysis of collagen XII stratified (tertiles) breast cancer patients from TCGA BRCA dataset ($n = 1067$). Univariate Cox Proportional Hazards Model hazard ratio (HR), HR 95% confidence interval and p value for collagen XII high vs low are indicated.

d Representative collagen XII IHC-stained TMA sections from triple-negative human breast cancer (TNBC) cohort⁶¹ of $n = 150$ patients (scale bar = $100 \mu\text{m}$). **e** Kaplan-Meier analysis of disease-specific survival and **f** distant recurrence from TNBC TMA sections, median-stratified by % stromal collagen XII staining ($n = 150$, 3 TMA cores per patient). Univariate Cox Proportional Hazards Model hazard ratio (HR), HR 95% confidence interval and p value for collagen XII high vs low are indicated. Source data are provided in the Source data file.

cancer-associated fibroblasts (CAFs) in this model (Fig. 5a-c). To further validate the relevance of this finding to the human setting, we also analysed scRNA-seq data from a cohort of 26 primary breast tumours⁶⁷ of the three major clinical breast cancer subtypes (ER+, TNBC and

HER2+) sampled from patients undergoing surgery. scRNA-seq analysis of 100,064 cells (Fig. 5d, e) further confirmed that CAFs are the major producers of collagen XII within primary human breast tumours (Fig. 5f).



The presence of CAFs within breast tumours has previously been associated with poor prognosis^{64,68}. To investigate the relationship between *COL12A1* and CAF presence in human tumours we assigned a CAF score to each breast tumour dataset from the TCGA cohort using the CAF marker genes identified in human scRNA-seq data from breast tumours above (Fig. 5f). Our analysis identified a strong positive correlation between *COL12A1* expression and CAF score, further

reinforcing CAFs as the major source of *COL12A1* in breast tumours (Supplementary Fig. 3A). In vitro and in vivo studies have demonstrated that CAFs can potentiate tumour progression^{64,68} and so in order to dissect the contribution of CAFs from that of collagen XII expression in the RNaseq data, we compared models of overall and progression-free survival with or without corrections for CAF score. Our analyses show that collagen XII expression, but not the CAF score,

Fig. 5 | Collagen XII is produced by cancer-associated fibroblasts in mouse and human breast tumours. **a** UMAP visualisation and clustering of 11,490 cells analysed by single-cell RNA-sequencing from Valdes-Mora et al.⁶⁶ of late-stage mammary tumours. **b** Feature plot and **c** Violin plot showing the log-normalised expression of *Coll2a1* over the UMAP structure and clusters as shown in **a** confirming the source of collagen XII expression. **d** UMAP visualisation and clustering of 100,064 cells analysed by single-cell RNA-Sequencing from the Breast Cancer Atlas⁶⁷, a study of 26 human primary breast cancers. Cell clusters are labelled by their major lineage annotations for B cells, Plasmablasts, T cells, Myeloid, Cancer epithelial, Normal epithelial, cancer-associated fibroblasts (CAFs), perivascular-like cells (PVL) and endothelial cells. **e** Feature plot and **f** Violin plot showing the log-normalised expression of *COL12A1* over the UMAP structure and

major lineage clusters as shown in **d**. **g** Western blot analysis of collagen XII secretion by normal mammary fibroblast (NF) and cancer-associated fibroblasts (CAFs) (*denotes non-specific band). GAPDH is loading control. Representative of $n = 3$ biologically independent experiments. **h** Quantification of *Coll2a1* mRNA expression ($n = 3$ biologically independent experiments **** $p < 0.0001$ one-way ANOVA with Tukey's multiple comparison test) and **i** collagen XII protein expression (representative of $n = 3$ biologically independent experiments) in collagen XII knockdown CAF lines generated using collagen XII short-hairpin mRNA scrambled control (shScr) and four collagen XII targeting constructs (shColXII#1–#4) (*denotes non-specific band). Vinculin is loading control. Source data are provided in the Source data file.

is significantly associated with overall and progression-free survival (Supplementary Table 8). Furthermore, we observe that the addition of the CAF score to the collagen XII survival model does not significantly improve the model (Supplementary Table 9), indicating that collagen XII expression is a predictor of outcome in these patients independently of the CAF score.

CAF-secreted collagen XII modulates collagen I organisation, tissue biomechanics and cancer cell invasion

To study the functional role of collagen XII in primary breast tumour progression in our model, we used CAFs and cancer cells isolated from primary PyMT tumours¹⁰. We confirmed collagen XII upregulation in these CAFs relative to normal murine mammary fibroblasts (NFs) (Fig. 5g and Supplementary Fig. 4A, B) in line with scRNA-seq data. We then generated stable collagen XII knockdown CAFs using four individual *COL12A1* short-hairpin RNA (shRNA) constructs (shColXII #1–#4). Following selection, stable knockdown was validated at both the mRNA (Fig. 5h) and protein (Fig. 5i and Supplementary Fig. 4C, D) level. Two independent lines (shColXII #1 and #3) showing significant collagen XII knockdown, plus control (shSCR)(with similar expression to the parental line) were used for functional validation of CAF-secreted collagen XII in primary breast cancer development and metastasis.

First, we sought to determine the effects that collagen XII knockdown had on de novo synthesis of cell-derived matrices (CDMs)^{69–71} (Fig. 6a). Imaging of collagen I in CDMs using SHG multiphoton imaging (Fig. 6b) showed no significant changes in the total amount of collagen I deposited, as measured by SHG peak signal intensity (Supplementary Fig. 5A). However the spatial patterning of the SHG signal, as measured by grey-level co-occurrence matrix (GLCM) analysis^{72–75} of the collagen fibres, showed a significant decrease in the uniformity of the collagen network following collagen XII knockdown (Fig. 6c). Importantly, these data indicate that knockdown of collagen XII expression does not alter the total abundance of fibrillar collagen, but does change fibrillar collagen ultrastructure. These data confirm that collagen XII is playing a role in determining spatial organisation of collagen I fibrils, and that the depletion of collagen XII in CAFs may act to normalise the matrix.

The 3D organisation and interaction of collagen family members are critically important in regulating the properties of the 3D microenvironment. In particular, changes in collagen I organisation are known to influence the biomechanical properties of tissues¹⁶, a parameter which has been shown to be increasingly important in solid tumour biology⁸. To determine the effect of collagen XII on collagen I organisation and biomechanics in 3D, shScr, shColXII#1 and shColXII#3 CAFs were seeded into and allowed to remodel, 3D collagen I organotypic matrices^{72–74,76} (Fig. 6d). Collagen XII knockdown did not alter CAF ability to contract the organotypic matrices (Supplementary Fig. 5B, C). However, unconfined compression analysis of the bulk biomechanical properties of the organotypic matrices, revealed that collagen XII depletion in CAFs led to a decrease in overall matrix stiffness (Fig. 6e).

Picrosirius red staining and polarised light birefringence imaging of the remodelled organotypic matrices (Fig. 6f) demonstrated a significant decrease in red (Fig. 6g) and yellow (Supplementary Fig. 5D) bundles as well as a corresponding increase in thin (green) fibrillar collagen bundles (Supplementary Fig. 5E) in the shColXII#1 and shColXII#3 matrices compared with shScr matrices. These data indicate that the depletion of collagen XII leads to a higher abundance of thinner collagen bundles. As collagen XII scaffolding underpins collagen I bundling, this is consistent with the known function of collagen XII in regulating collagen I architecture in healthy tissues^{30,43,44,77,78} and supports the observed decrease in biomechanical properties of these matrices.

Finally, we sought to determine whether the observed changes in the 3D organotypic matrices as a result of collagen XII depletion would affect cancer cell invasion in the 3D setting. PyMT cancer cells⁷⁹ were seeded onto, and allowed to invade into the CAF-remodelled organotypic matrices following removal of CAFs (see 'Methods') (Fig. 6h). While cancer cells readily invaded into shScr-CAF remodelled matrices, this ability was significantly reduced in collagen I matrices remodelled by the shColXII#1 and #3 CAFs (Fig. 6i, j). Similarly, a reduced cancer cell invasion was observed using the 4T1 murine breast cancer cell line (Supplementary Fig. 5F, G), further confirming that collagen XII regulation of collagen I is important in generating pro-invasive microenvironments in primary breast tumours.

To confirm that collagen XII expression by CAFs alters matrix biomechanics, fibrillar collagen architecture and cancer cell invasion, we employed a CRISPR-activation (dCas9-CRISPR-VPR) approach with guide RNAs (gRNAs) targeted to two different regions upstream of the collagen XII 5'UTR. This resulted in the generation of two CAF lines that overexpressed collagen XII (gRNA₁: ColXII-VPR₁; gRNA₂: ColXII-VPR₂) compared with CAFs expressing the non-targeting GFP control gRNA (eGFP-VPR). We confirmed that overexpression of collagen XII by CAFs did not affect the rate of collagen remodelling (Supplementary Fig. 5H) but did significantly increase the stiffness of CAF-remodelled organotypic matrices (Fig. 7a) and the proportion of mature, bundled fibrillar collagens (Fig. 7b and Supplementary Fig. 5I–K). Importantly, matrices remodelled by CAFs overexpressing collagen XII significantly increased the invasion of PyMT cancer cells (Fig. 7c, d). These data demonstrate that increasing collagen XII expression in CAFs exerts an opposite effect to collagen XII depletion, and confirm an important functional role for collagen XII in regulating fibrillar collagen architecture and matrix biomechanics in the tumour microenvironment and subsequently breast cancer cell invasion.

CAF-secreted collagen XII creates a permissive microenvironment that facilitates metastasis

Based on our in vitro data, we then investigated the role that collagen XII may be playing in breast cancer development and progression in vivo. To investigate the effects of CAF secreted collagen XII, we orthotopically co-implanted cancer cells with either shScr, shColXII#1 or shColXII #3 CAFs (1:3 ratio cancer cells to CAFs) into the fourth mammary fatpad (Fig. 8a). Following implantation, tumour growth was

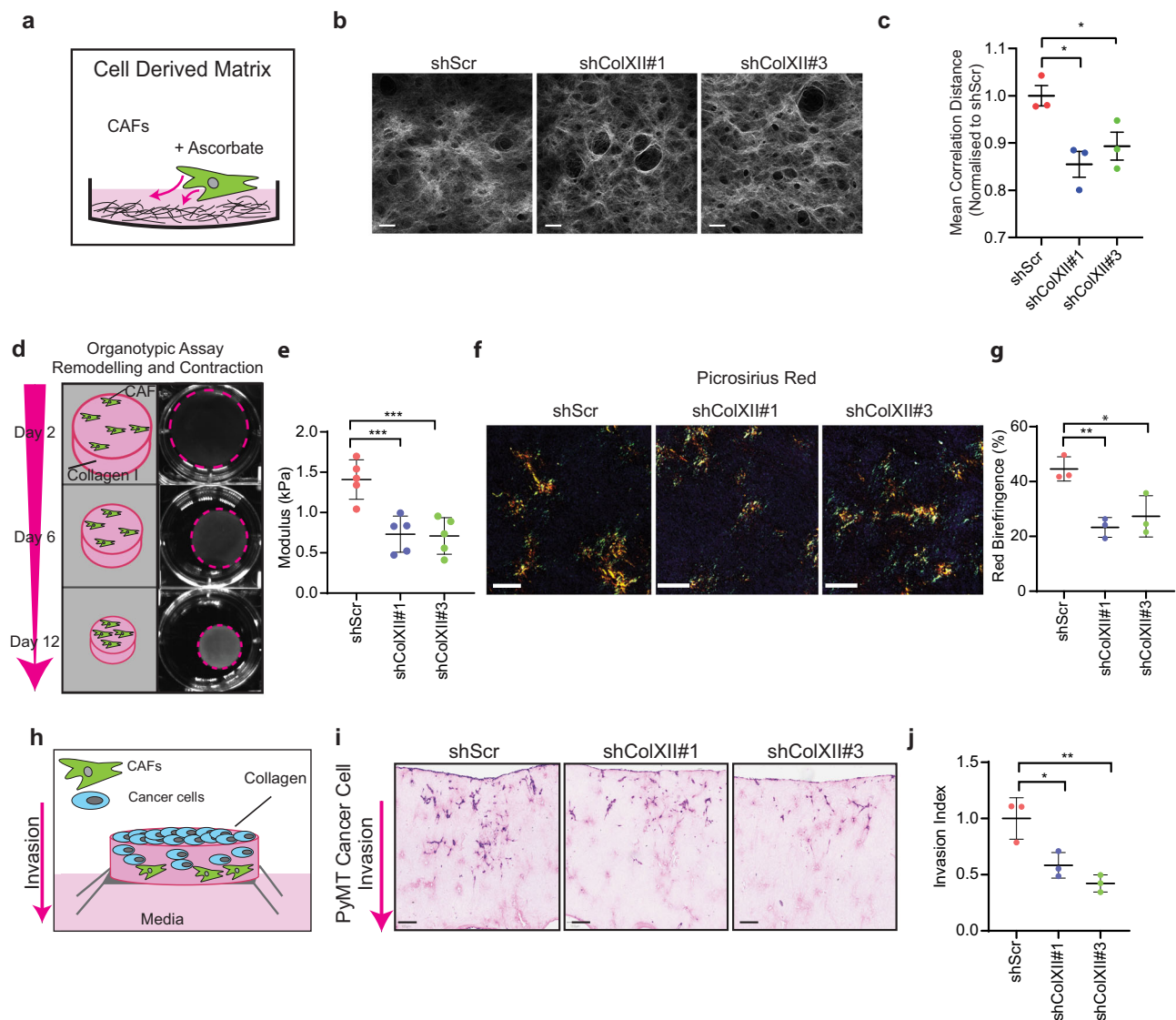


Fig. 6 | Collagen XII knockdown modulates fibrillar collagen architecture and inhibits cancer cell invasion. **a** Schematic representation of cell-derived extracellular matrix (CDM) secreted by cancer-associated fibroblasts (CAFs).

b Representative single-plane images of second harmonic generation (SHG) multi-photon signal of CDMs produced by CAF scrambled control (shScr) and two stable collagen XII knockdown lines (shColXII#1 and #3) from $n = 3$ biologically independent experiments. **c** Grey-level co-occurrence matrix (GLCM) analysis of mean correlation distance of collagen I structure in CAF-derived CDMs following collagen XII knockdown. Mean \pm SD, $n = 3$ biologically independent experiments, $*p < 0.05$, shScr vs. shColXII#1 $p = 0.14$; shScr vs. shColXII#3 $p = 0.049$; One-way ANOVA with Dunnett's multiple comparisons test. **d** Schematic representation of 3D organotypic matrix remodelling by fibroblasts (scale bar = 7 mm). Representative of $n = 5$ biologically independent experiments. **e** Bulk modulus (stiffness) of organotypic matrices remodelled by control (shScr) or collagen XII knockdown (shColXII#1, shColXII#3) CAFs at day 12 as measured by unconfined compression analysis. Mean \pm SD, $n = 5$ biologically independent experiments, $***p < 0.001$, shScr vs. shColXII#1 $p = 0.0006$; shScr vs. shColXII#3 $p = 0.0008$; one-way ANOVA with

Dunnett's multiple comparisons test. **f** Representative images of birefringence signal from picrosirius-red-stained organotypic matrices generated by control (shScr) or collagen XII knockdown (shColXII#1, shColXII#3) CAFs imaged under polarising light (scale bar = 100 μ m) from $n = 3$ biologically independent experiments. **g** Quantification of the red birefringence signal area for picrosirius-red-stained organotypic matrices remodelled by control (shScr) or collagen XII knockdown (shColXII#1, shColXII#3) CAFs. Mean \pm SD, $n = 3$ biologically independent experiments, $*p = 0.014$, $**p = 0.0054$, one-way ANOVA with a Dunnett's multiple comparisons test. **h** Schematic representation of experimental set-up of cancer cell invasion into CAF-remodelled organotypic matrices. **i** Representative histological images (H&E) ($n = 3$) of PyMT cancer cell invasion into organotypic matrices remodelled by control (shScr) or collagen XII knockdown (shColXII#1, shColXII#3) CAFs (scale bar = 100 μ m). **j** Quantification of PyMT cancer cell invasive index into organotypic matrices. Mean \pm SD, $n = 3$ biologically independent experiments, $*p = 0.020$, $**p = 0.0042$, One-way ANOVA with a Tukey's multiple comparison test. Source data are provided in the Source data file.

measured until a maximum endpoint tumour size of 520 mm³ (corresponding to 1 cm \times 1 cm) was reached. We observed no difference in time to endpoint between scrambled control (shScr) or collagen XII knockdown (shColXII #1 or #3) tumours, with all groups reaching maximum tumour volume within a similar timeframe (Supplementary Fig. 6A, B). Also, tumour weights at endpoint were not significantly different between scrambled control (shScr) and collagen XII knockdown (shColXII #1 or #3) tumours (Supplementary Fig. 6C). Building

on our in vitro data suggesting a role for collagen XII in tissue stiffness, we profiled freshly explanted tumours and found a significant decrease in bulk tumour stiffness, measured by unconfined compression analysis (Fig. 8b), in line with our organotypic studies (Fig. 7e).

In light of our in vitro data supporting a role for collagen XII in modulating cancer cell 3D invasion, we assessed metastatic burden within these mice. Macroscopic observation of H&E-stained sections of lung tissue (Supplementary Fig. 6D), one of the major sites of

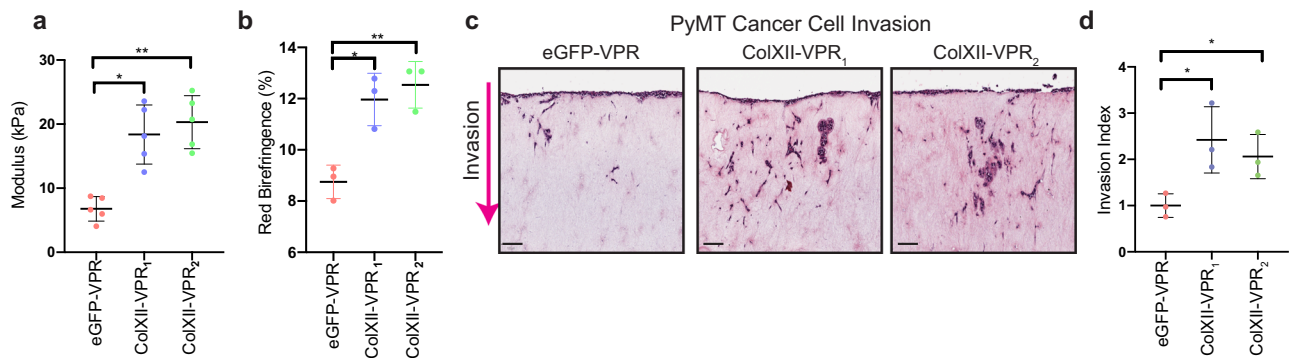


Fig. 7 | Collagen XII overexpression increases tumour stiffness and cancer cell invasion. **a** Bulk modulus (stiffness) of organotypic matrices at day 12 remodelled by control (eGFP-VPR) or collagen XII overexpressing (ColXII-VPR₁, ColXII-VPR₂) CAFs as measured by unconfined compression analysis. Mean \pm SD, $n = 5$ biologically independent experiments, $*p = 0.032$, $**p = 0.0075$, Kruskal–Wallis test with Dunn’s multiple comparison test. **b** Quantification of the red birefringence signal area for picosirius-red-stained organotypic matrices generated by control (eGFP-VPR) or collagen XII overexpressing (ColXII-VPR₁, ColXII-VPR₂) CAFs. Mean \pm SD, $n = 3$ biologically independent experiments, $*p = 0.034$, $**p = 0.0078$, one-way ANOVA with a Dunnett’s multiple comparisons test. **c** Representative histological

images (H&E) (from $n = 3$ biologically independent experiments) of cancer cell invasion into organotypic matrices remodelled by control (eGFP-VPR) or collagen XII overexpressing (ColXII-VPR₁, ColXII-VPR₂) CAFs (scale bar = 100 μ m).

d Quantification of cancer cell invasive index into organotypic matrices remodelled by control (eGFP-VPR) or collagen XII overexpressing (ColXII-VPR₁, ColXII-VPR₂) CAFs. Mean \pm SD, $n = 3$ biologically independent experiments, $*p < 0.05$, eGFP-VPR vs ColXII-VPR₁ $p = 0.032$; eGFP-VPR vs ColXII-VPR₂ $p = 0.028$; One-way ANOVA with a Dunnett’s multiple comparisons test. Source data are provided in the Source data file.

metastasis in breast cancer, revealed that approximately 50% of mice co-implanted with shColXII CAFs did not contain any detectable secondary lung tumours (Fig. 8c), compared to 100% of those co-implanted with shScr CAFs. Further analysis also revealed that in the 50% of mice with collagen XII knockdown CAFs that did exhibit lung metastases, the total number of metastases was significantly reduced (Fig. 8d), although there was no difference in the size of metastases present in mice bearing collagen XII knockdown tumours compared to control (Supplementary Fig. 6E). These data suggest that collagen XII plays a critical role in regulating matrix organisation that facilitates metastatic dissemination of tumour cells.

During primary tumour growth, the recruitment and co-option of host fibroblasts typically occur as has been described in ref. 64. IHC staining for collagen XII demonstrated a sustained significant decrease in collagen XII was still present within the knockdown CAF co-implanted tumours at endpoint compared to scrambled control (Fig. 8e, f and Supplementary Fig. 6F), which was confirmed by western blotting (Fig. 8g). To ensure that this was not due to overall lower numbers of CAFs in the collagen XII knockdown tumours, IHC staining for the CAF marker α SMA¹⁰ showed no differences in CAF presence within the different tumour groups at endpoint (Supplementary Fig. 6G, H). Thus these data confirm that the lower collagen XII expression in knockdown tumours was not due to a lower overall CAF presence in these tumours, and suggest that recruitment of local fibroblast populations cannot restore collagen XII levels. Furthermore, in line with our in vitro findings, SHG multi-photon imaging of collagen I organisation (Fig. 8e) showed a significant decrease in collagen I bundle thickness (Fig. 8h), correlating with the observed decrease in tumours stiffness.

During tumour progression, fibroblasts within the tumour are known to exhibit increased activation of cytoskeletal contractile machinery and their regulators, including pMLC2, which plays a central role in a positive feedback loop further activating these fibroblasts and amplifying matrix remodelling^{10,80}. This matrix remodelling leads to increases in tumour stiffness and the generation of microenvironments that promote tumour cell invasion. Examining the level of MLC2 phosphorylation in primary tumours at endpoint revealed strong pMLC2 staining in the stroma of these tumours. Furthermore, pMLC2 levels were significantly decreased in collagen XII knockdown tumours compared with control tumours (Fig. 8i and Supplementary Fig. 6i), consistent with reduced primary tumour stiffness and subsequently

decreased metastatic potential of cancer cells in an environment with low collagen XII expression. These data confirm that collagen XII depletion in CAFs leads to changes in collagen I bundling and tumour stiffness within the developing tumours, that likely leads to an overall lower level of stromal activation and the abrogation of a cancer cell invasion permissive microenvironment.

Finally, to further validate the functional importance of Collagen XII expression in metastatic dissemination, we orthotopically co-implanted cancer cells with CAFs overexpressing collagen XII (ColXII-VPR₁, ColXII-VPR₂) or control CAFs (eGFP-VPR), into the fourth mammary fatpad (1:3 ratio cancer cells to CAFs) similar to the knockdown experiments. Overexpression of collagen XII within the primary tumour was associated with a small increase in primary tumour progression (Supplementary Fig. 7A–C). Consistent with the data from the knockdown study, collagen XII overexpression by CAFs resulted in stiffer tumours (Fig. 9a), concordant with the role of this FACIT collagen in regulating tumour biomechanics. Importantly, overexpression of collagen XII in tumours also led to a higher number of lung metastases compared with control tumours (Fig. 9b), although did not significantly affect the number of mice with lung metastases (Supplementary Fig. 7D, E), nor did it affect the size of metastases present (Supplementary Fig. 7F). Finally, we confirmed that at endpoint, there was a sustained increase in collagen XII expression within the tumour microenvironment as measured by both IHC (Fig. 9c and Supplementary Fig. 7G) and western blotting (Supplementary Fig. 7H, I). Consistent with the knock-down study, there was a trend towards increased pMLC2 staining in tumours overexpressing collagen XII (Fig. 9d and Supplementary Fig. 7J), reinforcing collagen XII-induced biomechanical changes in the primary breast tumour microenvironment that support cancer cell metastasis.

Overall, these complementary knockdown and overexpression in vitro and in vivo models support the hypothesis that CAF-secreted collagen XII at the primary site modulates collagen I organisation and tumour biomechanics, thereby creating a metastasis permissive tumour microenvironment.

Discussion

The development and progression of many solid tumours is accompanied by a desmoplastic response whereby the deposition and remodelling of the tumour matrix lead to significant changes in the biochemistry, biomechanics, architecture, and topography of the

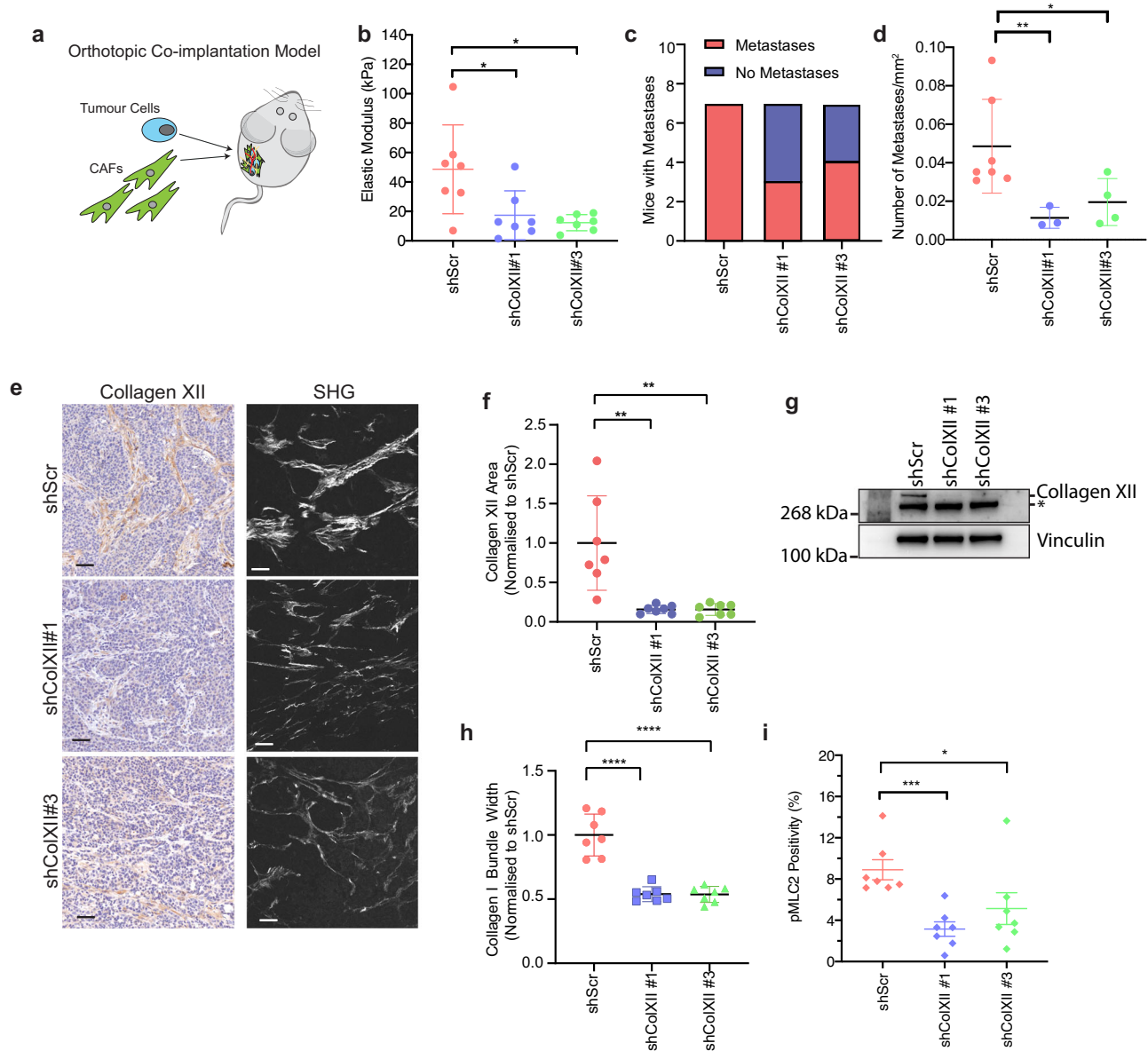


Fig. 8 | Collagen XII knockdown reduces metastasis in vivo. **a** Schematic representation of orthotopic co-implantation of cancer cells and CAFs (knockdown study: shScr and shColXII; overexpression study eGFP-VPR and ColXII-VPR) injected in a 1:3 ratio respectively. **b** Bulk modulus as determined by unconfined compression analysis of excised primary tumours generated by co-implantation of cancer cells with control (shScr) and collagen XII knockdown (shColXII#1, shColXII#3) CAFs. Mean \pm SD, $n = 7$ biologically independent samples, $*p < 0.05$, shScr vs shColXII#1 $p = 0.036$, shScr vs. shColXII#3 $p = 0.045$; Kruskal–Wallis test with Dunn’s multiple comparisons test. **c** Quantitation of the number of mice with or without observed metastases in H&E sections in the knockdown study ($n = 7$ mice per group). **d** Quantification of the mean number of metastases normalised to the total lung area in mice with metastases present in the knockdown study ($n = 7$ shScr, $n = 4$ shColXII#1 and $n = 3$ shColXII#3 mice with metastases present). $*p = 0.0278$, $**p = 0.0062$ two-sided t test with Welch’s correction. Data are presented as mean \pm SD. **e** Representative images of (left) collagen XII IHC (scale bar = $50 \mu\text{m}$)

and (right) multi-photon second harmonic generation signal (scale bar = $40 \mu\text{m}$) acquired from $n = 7$ primary tumours in the knockdown study. **f** Quantification of collagen XII-positive area in primary tumours from the knockdown study ($n = 7$ mice per group). $**p < 0.01$ One-way ANOVA with a Dunnett’s multiple comparisons test. Data are presented as mean \pm SD. **g** Representative western blot of collagen XII expression in primary tumours at endpoint in the knockdown study of $n = 2$ biologically independent experiments. *Denotes non-specific band. Vinculin is a loading control. **h** Quantification of collagen I fibre bundle width in primary tumours from the knockdown study. Data are presented as mean \pm SD, $n = 7$ biologically independent experiments, $****p < 0.0001$, one-way ANOVA with Dunnett’s multiple comparisons test. **i** Quantification of pMLC2 staining positivity (% of tumour area) in primary tumours. Data are presented as mean \pm SD, $n = 7$ mice per group; 3 fields of view per tumour, $*p = 0.018$, $***p = 0.0006$, two-sided Mann–Whitney test. Source data are provided in the Source data file.

tumour microenvironment^{41,81,82}. It is well established that the ECM is a salient feature of solid tumours and can play both a pro-tumourigenic^{8–13}, as well as anti-tumourigenic role^{1,5,14,15}. In particular, the total and relative amounts of each matrix component, as well as the specific 3D supramolecular assembly and organisation are critically important^{1,5}. It is clear that increased matrix deposition alone is not always a robust indicator of patient outcome³, and given the

paradoxical role of the matrix in cancer, recent work has focussed on uncovering the critical elements which regulate matrix assembly and organisation, and the role that they play in solid tumour progression. Many studies have also implicated CAFs⁶⁴ as one of the key architects of matrix deposition and remodelling in solid tumours⁸³.

While single timepoint ECM profiling studies have revealed critical insight into how the matrix differs between tumour and non-

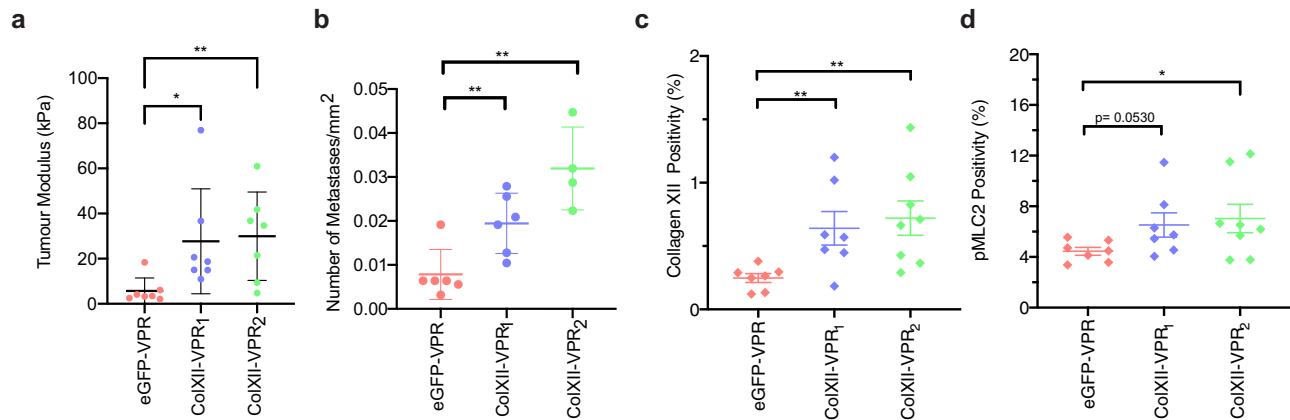


Fig. 9 | Collagen XII overexpression promotes metastasis in vivo. **a** Bulk modulus as determined by unconstrained compression analysis of excised primary tumours generated by co-implantation of cancer cells with control (eGFP-VPR) and collagen XII overexpressing (ColXII-VPR₁, ColXII-VPR₂) CAFs. Data are presented as mean \pm SD, $n = 7$ eGFP-VPR mice, $n = 7$ ColXII-VPR₁ and $n = 8$ ColXII-VPR₂ mice, $*p = 0.017$, $**p = 0.0066$, Kruskal–Wallis test with Dunn’s multiple comparisons test. **b** Quantification of mean number of metastases normalised to the total lung area in mice with metastases present in the overexpression study ($n = 6$ eGFP-VPR mice, $n = 6$ ColXII-VPR₁ and $n = 4$ ColXII-VPR₂ mice with metastases present). Data are presented as mean \pm SD. eGFP-VPR vs. ColXII-VPR₁ $p = 0.0087$; eGFP-VPR vs. ColXII-

VPR₂ $p = 0.0048$, two-sided Mann–Whitney U -test. **c** Quantification of collagen XII positive area in primary tumours from the overexpression study. $n = 7$ eGFP-VPR mice, $n = 7$ ColXII-VPR₁ and $n = 8$ ColXII-VPR₂ mice. $**p < 0.01$, eGFP-VPR vs. ColXII-VPR₁ $p = 0.0098$; eGFP-VPR vs. ColXII-VPR₂ $p = 0.025$, two-sided Mann–Whitney U -test. One-way ANOVA with a Dunnett’s multiple comparisons test. Data are presented as mean \pm SD. **d** Quantification of pMLC2 positive area in primary tumours from the overexpression study. $n = 7$ eGFP-VPR mice, $n = 7$ ColXII-VPR₁ and $n = 8$ ColXII-VPR₂ mice. $*p = 0.04$, two-sided Mann–Whitney U -test. Data are presented as mean \pm SD. Source data are provided in the Source data file.

tumour tissues^{31,33,35}, matrix remodelling is a highly dynamic process. Insight into how the matrix changes throughout tumour progression will reveal the individual and collective contribution of key components at critical clinically important stages of tumour progression, such as metastatic dissemination. Here, using matrisome-enriched temporal proteomic profiling, we interrogated the changing tumour matrisome during breast cancer progression in a spontaneous autochthonous immunocompetent model of breast cancer. This identified 4 key matrisomal clusters with different temporal profiles that likely play important and different roles during progression. For example, early upregulation of cluster 2 proteins such as SPPI may represent a specific matrix remodelling programme associated with initial fibroblast trans-differentiation into CAFs that accelerates tumour progression during the early stages of transformation⁵⁰. Conversely, cluster 1 proteins such as fibulin 5 have also been implicated in promoting cancer cell invasion in other cancer types⁸⁴ and therefore may represent a temporal profile that supports metastatic dissemination.

In particular, we pinpoint the cluster 1 protein collagen type XII as a critical component that is upregulated in tumours compared to healthy tissue, and which is also increasingly deposited over time as tumours progress from hyperplastic lesions to adenoma and through to metastatic adenocarcinoma⁸⁵. Through cross-comparison of our proteomic profiling with single-cell transcriptomic analysis of tumours from matched animal models and patient tumours, we revealed the source of collagen XII as CAFs. We show that collagen XII, a FACIT collagen, is predominantly secreted by CAFs and is important in regulating the spatial organisation of collagen I within the primary tumour microenvironment. Collagen XII is the largest member of the FACIT collagens³⁰ and is known to bind to type I collagen fibrils to regulate their organisation and contribute to tissue biomechanics in the cornea, skeletal muscle and tendon^{30,43,44,86}. Emerging evidence is now revealing dysregulation of this ECM component in several cancer types. Recently, collagen XII expression has been correlated with high mammographic density in women, a known risk factor for poor outcome in breast cancer²⁰. Collagen XII has also previously been associated with poor clinical outcomes in gastric cancer⁵⁵; however, to date, a functional role for collagen XII in breast cancer metastasis had not been reported.

Analysis of both human and murine single-cell transcriptomic data identified CAFs as a major source of collagen XII in the tumour microenvironment, concordant with previous reports of high expression of collagen XII in colorectal carcinoma myofibroblasts in vitro⁴⁷. Importantly, our data indicate that collagen XII expression levels in the primary tumour are not simply a marker of CAF presence within the tumour, but that this CAF-secreted collagen XII plays a functional role in creating an invasion-permissive 3D microenvironment that supports metastatic dissemination of breast cancer cells. Stable manipulation (knockdown or overexpression) of collagen XII expression in CAFs in our co-implantation in vivo orthotopic model significantly altered collagen I organisation and biomechanics of the tumour tissue. This subsequently leads to a significant alteration of dissemination to the lung, with increases in collagen XII accompanying increased metastasis.

Since collagen XII levels significantly affected metastasis number but not time to endpoint in the knockdown study, our data suggest that elevated collagen XII levels in the primary tumour facilitate cancer cells in disseminating to the secondary tissues such as the lung. This increase in metastatic potential may arise from either more cancer cells disseminating from the primary tumour and/or an increased survival advantage that facilitates arrival at the lung. Our in vitro data suggests that collagen XII levels at the primary tumour site modulate the invasiveness of cancer cells, lending support to the former mechanism of metastasis potentiation. Importantly, this suggests that a future intervention targeting collagen XII-cancer cell crosstalk may disrupt the pro-invasive effects of a collagen XII-rich tumour microenvironment. However, since an almost complete depletion of collagen XII did not abrogate metastasis in general, this reinforces the notion that collagen XII is likely one of many elements within the tumour microenvironment affecting dissemination.

Importantly, our data also indicate that in the primary tumour microenvironment, collagen XII regulates collagen I organisation. The organisation of collagen I fibrils, and the associated changes in tissue biomechanics, have been widely implicated in promoting cancer cell invasion and metastasis^{2,8,18}, largely through mechanotransduction mechanisms. Importantly, increased stiffness also acts on normal fibroblasts to drive them towards a myofibroblast phenotype that accelerates further tumour-promoting matrix remodelling, a phenotype

supported by our pMLC2 data^{10,87}. Emerging roles for collagen I architecture and tumour biomechanics in immunosurveillance^{88,89} suggests that the effects of collagen XII on these parameters may have broader implications for immunological constraints on tumour progression. Overall, collagen XII expression has the potential to reinforce matrix remodelling leading to the generation of an environment that facilitates metastatic dissemination.

In analysing expression levels of collagen XII in patient cohorts we found that high expression of collagen XII is significantly associated with poor progression-free and overall survival in breast cancer patients. Importantly, collagen XII was found to be a significant predictor of progression-free survival in early-stage (stage I–II) patients, a population in which overt metastatic dissemination is not detected at the time of diagnosis. Within this population, the prediction of recurrent disease remains a considerable clinical challenge. Perhaps surprisingly, collagen XII expression was the strongest predictor of progression-free survival in this group compared to other known clinicodemographic indicators of risk. Supported by our *in vitro* and *in vivo* data, this suggests that in breast cancer, collagen XII may be playing a key role in early dissemination to secondary sites. With further development and validation in independent cohorts, and importantly in a large prospective study, collagen XII expression in the primary tumour stroma may find utility as a biomarker of early dissemination and therefore high risk of recurrence in early-stage breast cancer patients.

Finally, our data also show that matrisomal profiling is a powerful method to reveal dynamic changes in the matrix and understanding these changes may present new underexplored stromal targets that play important roles at critical stages of tumour progression. In recent years, there has been a marked increase in interest in the role of the matrix in defining the properties of the tumour microenvironment, and our findings support the emerging hypothesis that future therapeutic approaches to ‘normalise’ or re-engineer the tumour stroma may offer the potential for significant translational impact in improving patient survival.

Methods

Study design

All animal research was conducted according to protocols approved by the St Vincent’s Precinct and Garvan Institute Animal Ethics Committee (protocol # ARA19_08). All research involving biobanked human TMA biospecimens and data was covered by the Royal Prince Alfred Hospital Human Ethics Review Committee Approval (X14-0241). Informed consent had been provided for the study and compensation was not provided. The number of mice used in each experiment is detailed in corresponding figure legends. Wild-type (wt) or polyoma middle-T antigen (PyMT) mammary tissues were harvested at early (8–10 weeks), mid (11–13 weeks) or late (14–16 weeks) timepoints. In the tumour setting, this corresponds to hyperplasia, adenoma, followed by carcinoma⁹⁰. Tissue biomechanical properties were quantified using unconfined compression analysis on freshly collected samples. Tumour and normal tissue were then utilised in either histological or proteomic mass spectrometry workflows. LC-MS/MS was performed on 5 independent biological replicates (with the exception of mid-stage tumour which consisted of 4 biological replicates) in a single proteomic run. CDMs and organotypic contraction and invasion assays were performed in three independent biological repeats with three technical replicates. Picosirius red, IHC, SHG, GLCM, orientation and fibre bundle width analyses were performed on three regions of interest in CDMs and organotypic matrices and five regions of interest in animal tissues.

Animal studies and ethics

Female FVB/n mice with or without the polyoma middle-T antigen (PyMT) transgene under the mouse mammary tumour virus (MMTV)

promoter were obtained in house⁶⁶ in conventional animal facilities in line with the Australian code of practice for the care and use of animals for scientific purposes, including standard ambient temperature, humidity and dark/light cycles. Genotyping of genetically engineered mouse models was performed by Garvan Molecular Genetics (GMG) facility (Sydney, Australia). For co-implantation studies, female FVB/n mice aged 10–12 weeks were used as detailed below in “Orthotopic *in vivo* study”. For the knockdown study $n=7$ mice were used per group. For the overexpression study $n=8$ mice were used per group; however, one mouse in each of the eGFP-VPR and ColXII-VPR₁ groups did not develop tumours so was excluded from further analysis. All animal work was carried out in accordance with protocols approved by the St Vincent’s Precinct and Garvan Institute Animal Ethics Committee (protocol # ARA19_08).

Cell lines and cell culture

wt mammary fibroblasts (NFs) derived from wt FVB/n mice (NFs), or cancer associated fibroblasts from transgenic PyMT FVB/n mice (CAFs) were a kind gift from Fernando Calvo¹⁰. Fibroblast lines were maintained in Dulbecco’s modified Eagle media (DMEM) supplemented with 10% FBS, 1% ITS (insulin–transferrin–selenium) and 1% penicillin/streptomycin. PyMT 20065 cancer cells were derived with the support of Dr Karen Blyth at the CRUK Beatson Institute in Glasgow through the SEARCHbreast initiative (<https://searchbreast.org/>), a resource to facilitate sharing of archived material derived from *in vivo* breast cancer models. PyMT 20065 cancer cells were maintained in DMEM supplemented with 10% FBS, 1% penicillin/streptomycin, 5 µg/mL insulin, 10 ng/mL epidermal growth factor and 10 ng/mL Cholera Toxin A as previously described⁷⁹. All cells were kept at 37 °C in 20% O₂ and 5% CO₂. Cells were routinely tested and confirmed negative for mycoplasma.

Proteomics

Proteomic sample preparation. To enrich matrix proteins, tissue was incubated for 16 h with shaking at room temperature in 0.5% sodium deoxycholate (SDC). Samples were centrifuged at 1000 × *g* for 2 min and the supernatant was discarded. Pellets containing decellularized proteins were washed briefly with 0.5% SDC, centrifuged at 1000 × *g* for 2 min and the supernatant discarded. Proteins were resuspended in 1% sodium dodecyl sulfate (SDS) in 100 mM Tris, pH 8.5 and solubilised with 2 × 20 s tip-probe sonication. Protein was quantified using a BCA and normalised to 20 µg/100 µL of 1% SDS in 100 mM Tris, pH 8.5. The samples were reduced and alkylated in a final concentration of 10 mM Tris(2-carboxyethyl)phosphine and 40 mM 2-Chloroacetamide, respectively for 5 min at 45 °C. Peptides were prepared using a modified single-pot, solid-phase-enhanced sample preparation as previously described⁹¹. Briefly, samples were diluted to 50% ethanol and incubated with 1:1 mixture of hydrophilic:hydrophobic Seramag Speedbead carboxyl magnetic beads (GE Life Sciences) for 8 min at room temperature. The supernatant was removed, and the beads were washed three times with 80% ethanol. The beads were then resuspended in 10% trifluoroethanol in 100 mM Tris-HCl, pH 7.5 and digested with 0.4 µg of sequencing-grade trypsin (Sigma) and 0.4 µg of sequencing grade LysC (Wako, Japan) overnight with shaking at 37 °C. Trypsin digestion was halted by acidification to 1% trifluoroacetic acid (TFA) and purified by styrenedivinylbenzene-reverse phase sulfonate (SDB-RPS) microcolumns. The columns were washed with 99% isopropanol containing 1% TFA followed by 5% acetonitrile containing 0.2% TFA and peptides eluted with 80% acetonitrile containing 1% ammonium hydroxide and dried by vacuum centrifugation. Peptides were resuspended in 2% acetonitrile containing 0.1% TFA and stored at –20 °C prior to running.

Mass spectrometry acquisition and data processing. Prepared peptide samples were analysed using a Dionex nanoUHPLC coupled to

a Q-Exactive HF-X in positive polarity mode using XCalibur. Peptides were separated on 20 cm × 100 μm column constructed in-house with an integrated emitter and packed with 1.9 μm C18AQ particles (Dr. Maisch, Germany). One microgram of the peptide was injected and eluted over a linear gradient of 3–35% Buffer B over 60 min at 60 °C with a flow rate of 800 nl/min (Buffer A = 0.1% formic acid; Buffer B = 80% acetonitrile, 0.1% formic acid). The mass spectrometer was operated in data-independent acquisition (DIA) using identical settings and variable-sized isolation windows as previously described⁹². DIA data were analysed in Spectronaut Pulsar X using library-free searching. Peptide quantification was performed at MS2 level using 3–6 fragment ions, which included automated interference fragment ion removal as previously described⁹². MS1 mass tolerance was set to 20 ppm and the MS/MS fragment mass tolerance was set to 0.02 Da. Dynamic mass MS1 and MS2 mass tolerance were enabled and local (non-linear) regression was performed for retention time calibration. A dynamic extracted ion chromatogram window size was performed. The minimum peptide length was set to 7 amino acids with specific trypsin cleavage and search criteria included oxidation of methionine and protein N-terminal acetylation set as variable modifications and carbamidomethylation set as a fixed modification. Data were searched against the mouse UniProt (June 2018; 95,128 entries including isoforms) and filtered to 1% FDR at the peptide and protein level (Q-value cut-off <0.01). Peptide quantification was performed using 3–6 fragment ions, and protein quantification was performed with weighted peptide median values. Perseus software was utilised for statistical analysis. Proteins with <70% presence across all samples were filtered in order to yield robustly identified proteins. For a targeted matrix analysis, proteomics data were annotated for matrisome proteins³⁵, and non-matrisome proteins were filtered. Differentially abundant proteins were calculated using *t* tests with multiple hypothesis testing using the Benjamini–Hochberg adjustment. Significant differences were defined with an adjusted *p* and *q* < 0.05. Heatmaps were generated using log₂ transformed, median-centred *z*-score of protein abundance values, from proteins present in at least 70% of all samples (*n* = 4 or 5). Volcano plot analyses were generated using log₂ transformed and median-centred protein abundance values of proteins present in at least 70% of all samples (*n* = 4 or 5). Principal component analysis biplots were generated using log₂ transformed and median-centred protein abundance values, of proteins in all samples (*n* = 4 or 5).

Unconfined compression analysis

Biomechanical testing of tissues was performed using the TA Instruments Discovery Hybrid Rheometer-3 (DHR3) with TRIOS Data acquisition software. Fresh tumours and mammary glands were isolated at defined timepoints (early [8–10 weeks], mid [11–13 weeks], or late [14–16 weeks]) and immediately subjected to biomechanical testing. Organotypic matrices were profiled after 12 days of remodelling. Compressive elastic modulus was measured by applying a constant linear pressure of 2 μm/s for animal tissues and 10 μm/s for organotypic matrices. The data were acquired and a stress/strain curve for each tissue was obtained. The compressive elastic modulus was obtained from the slope of the linear viscoelastic region of the stress/strain curve accounting for material surface area.

Immunoblotting

Samples were lysed in protein extraction buffer (50 mM HEPES, 1% Triton-X-100, 0.5% sodium deoxycholate, 0.1% SDS, 0.5 mM EDTA, 50 mM NaF, and protease cocktail inhibitor (Roche)). Proteins were separated by either 4–12% acrylamide Bis-Tris or 3–8% tris-acetate gel electrophoresis (as detailed in individual figure legends), and were transferred onto a PVDF membrane (Immobilon-P, Millipore), blocked with 5% milk, incubated with primary antibodies (COL12A1 1:10,000 Abcam ab121304; B-tubulin 1:10,000 Cell signalling #2146, Vinculin

1:10,000 Sigma V9131, GAPDH 1:10,000 Cell Signalling #2118, clone 14C10) overnight at 4 °C, and probed with HRP-conjugated secondary antibodies (#NA931, GE Healthcare, 1:5000 diluted in 1% milk). Signal was detected using ECL reagent and imaged (Fusion FX, Vilber). Uncropped blots are included in the Supplementary Information File and Source data file.

Histopathology and immunohistochemistry

Tissues were fixed in 10% buffered formalin and embedded in paraffin and sections were taken at 4 μm. Sections were deparaffinised in xylene and rehydrated using a series of graded ethanol washes. Staining was performed using the Bond RX Autostainer (Leica). Heat-Induced Epitope Retrieval was performed in an EDTA buffer pH 9 (Leica, AR9640) for 40 min at 93 °C. Slides were quenched in Peroxide Block (Leica, DS9800). Primary antibodies were used as follows: COL12A1 (1:150) Abcam ab121304; αSMA (1:150) ab5694 Abcam; pMLC2 (1:100) Cell Signalling Technologies #3671. Staining used the Leica Bond Polymer Refine Detection Kit (Leica, DS9800) as per the manufacturer's instructions. Skeletal muscle was used as a positive control tissue for optimisation of COL12A1 staining. Staining was visualised using Diaminobenzidine (DAB). H&E staining and counterstaining were performed on the Leica ST5010 Autostainer XL (Leica). Quantification of the DAB area was performed in ImageJ (v2.3.501). Whole stained tissue sections were imaged using an Aperio slide scanner.

Picrosirius red staining and quantification

Picrosirius red staining was performed as previously described⁷³. The area of total collagen stained by picrosirius red was quantified in ImageJ⁹³ software using an in-house script as previously published⁷³. Briefly, 4 μm sections of fixed samples were deparaffinised, rehydrated, and stained with 0.1% picrosirius red (Polysciences) for fibrillar collagen according to the manufacturer's instructions. Imaging was performed on Leica DM 6000 (Power Mosaic) at ×20 magnification for animal tissues and ×40 magnification for organotypic collagen matrices. Quantitative measurements of fibrillar collagen signal (Red) were carried out. For each image, Hue-Saturation-Balance (HSB) thresholding was applied, where $200 \geq H \leq 240 \mid 150 \geq S \leq 255 \mid 0 \geq B \leq 255$ was used. The relative area (as a % of total tissue area) was then calculated.

SHG imaging

SHG imaging was performed on formalin-fixed paraffin-embedded unstained 4 μm sections as previously described⁷³. Collagen I SHG signal was imaged using a 25 × 0.95 NA water objective on an inverted Leica DMI 6000 SP8 confocal microscope. Excitation was achieved using a Ti:Sapphire femtosecond laser (Coherent Chameleon Ultra II) at 80 MHz and tuned to a wavelength of 920 nm. The intensity was recorded with RLD HyD detectors at 460/20 nm. For each subject, five representative regions of interest of 1024 μm × 1024 μm were imaged over a depth of 25.2 μm with a *z*-step size of 2.52 μm. The intensity of the SHG signal was quantified using ImageJ.

GLCM analysis

Collagen fibre organisation was assessed using GLCM. GLCM analysis was performed using previously published code⁷³ as previously described^{72,73}. In-house ImageJ and MATLAB scripts are available via GitHub (<https://github.com/TCox-Lab>). This method quantifies the correlation of SHG signal throughout the matrix, characterising the texture of a sample. The correlation graphs represent the similarity between pixels, where a higher mean correlation demonstrates more organisation within the collagen network, whilst a low correlation demonstrates a less ordered matrix. GLCM analysis was performed in MATLAB (Mathworks, vR2020), using three representative single-plane SHG images. The average texture parameter for each image was calculated using the looped operation of the plug-in and for 0°, 90°, 180° and 270° directions. Normalised texture parameters were

calculated for each image, and the mean correlation along with the SEM was imported and plotted in the GraphPad software.

Collagen fibre bundle width analysis

To measure the width of collagen I fibre bundles, five representative single-plane SHG images were analysed using the Profile Plot function in ImageJ. Outputs generate a signal intensity vs. pixel position histogram, where peak width corresponds to fibre bundle diameter. Within each image, the average fibre bundle width was calculated by analysing the width of collagen fibres in each section of a 3 × 3 grid applied to each ROI.

Orientation analysis

To quantify collagen I fibre orientation as previously described^{31,72,73,94}, an in-house ImageJ script was used. Briefly, structure tensors are utilised to derive the local orientation and isotropic properties of pixels. Tensors were evaluated for each pixel of an input image by computing the continuous spatial derivatives in the *x* and *y* directions using a cubic B-spline interpolation to obtain the local predominant orientation. A hue-saturation-brightness colour-coded map image output is used to indicate the angle of the oriented collagen I within the image. Orientation distribution peaks were then aligned. Distribution shapes denote the degree of linearisation within the image, where wide and broad shapes suggested little coherency in alignment, and narrow peaks imply aligned structures. The ratio between the peak and baseline of each curve was calculated to give the degree of alignment⁹⁵.

Analysis of COL12A1 expression in patient datasets

'The Cancer Genome Atlas' (TCGA). All gene expression analysis was performed in R (v3.6.1). To assess the relative expression of *COL12A1* in non-tumour and tumour tissue, RNAseq expression values were downloaded from GDAC firehose. The RNAseq data were filtered for lowly expressed genes before being normalised using EdgeR and Limma^{96,97}. TCGA RNAseq from the pan-cancer dataset⁶⁰ (RSEM normalised), corresponding clinicodemographic information was obtained from GDAC firehose and survival information was obtained from⁹⁸. *COL12A1* expression was stratified into tertiles, with the upper tertile defined as 'high' *COL12A1* expression. Multivariate analyses of *COL12A1* together with clinical covariates included age, stage, a binary variable indicating whether positive lymph nodes were detected and receptor subtype status (Luminal A: ER or PR positive + Her2 negative; Luminal B: ER or PR positive + Her2 negative; Her2: ER and PR negative + Her2 positive; Triple Negative: ER, PR and Her2 negative) as covariates.

CAF contribution to survival in the TCGA cohort. A human CAF signature was obtained as the CAF marker genes reported by Wu et al. (2021)⁶⁷. *COL12A1* was removed from this signature to avoid its confounding effect on examining the relationship between *COL12A1* and the CAF signature. A CAF score was generated for each TCGA sample using Gene Set Variation Analysis⁹⁹ and default parameters. The association of the CAF score with overall and progression-free survival was assessed using cox proportional hazards modelling (coxph function of the survminer package in R). Model comparisons were made using the *anova* function of the survival package.

Tumour microarrays. Tumour microarrays (TMAs) were constructed from *n* = 150 patients with triple-negative breast cancer (TNBC) as previously described⁶¹. The TMAs contained at least three cores taken from different areas of each patient's tumour. TMAs were stained by IHC for collagen XII. First, 4 µm sections were deparaffinized in xylene and rehydrated using a series of graded ethanol washes. Staining was performed using the Bond RX Autostainer (Leica). Heat-Induced Epitope Retrieval was performed in an EDTA buffer pH 9 (Leica, AR9640) for 40 min at 93 °C. Slides were quenched in Peroxide Block (Leica, DS9800). The slides were incubated with primary antibodies (COL12A1

1:200 Abcam ab121304) and the Leica Bond Polymer Refine Detection Kit (Leica, DS9800) was used for detection in combination with Diaminobenzidine (DAB). H&E staining and counterstaining were performed on the Leica ST5010 Autostainer XL (Leica). Stained sections were imaged using an Aperio slide scanner. To assess the association of *COL12A1* with survival and clinicodemographic features, blinded scoring of *COL12A1* IHC staining in stromal compartments of the tumour tissue was carried out by a registered pathologist (Prof Sandra O'Toole, Department of Tissue Pathology and Diagnostic Oncology, Royal Prince Alfred Hospital and NSW Health Pathology, Sydney, NSW, Australia). The maximum *COL12A1* score was calculated across each of the three cores per patient and patients were subsequently stratified into high and low groups by the median. The association of *COL12A1* stromal scores with disease-specific and distant recurrence was assessed using univariate cox proportional hazards modelling and Kaplan-Meier analysis (Survminer and Survival packages). Univariate Cox proportional hazards modelling of clinical variables identified that Age and lymphatic invasion were significantly associated with measures of survival and therefore, these parameters were included as terms in a multivariate model with the *COL12A1* scores in order to assess the independent prognostic importance of *COL12A1* expression.

Single-cell RNA sequencing

PyMT mammary tumour. Single-cell RNAseq data of PyMT tumours were obtained from the Valdes-Mora et al. 2021 dataset (GSE158677)⁶⁶ selecting only the MMTV-PyMT genotype containing 11,490 cells from five different tumours. This subset was analysed using the Seurat (v3.2) package¹⁰⁰, establishing QC thresholds for cell calling of <5% mitochondrial to nuclear gene content and <8000 molecules/cell to exclude doublets. Downstream analysis was performed according to ref. 100, including linear dimensional reduction (PCA), building a K-nearest neighbour (KNN) graph using 30 principal components "FindNeighbors" and default clustering parameters "FindClusters". Data visualisation was performed using non-linear dimensional reduction UMAP¹⁰¹.

Human Breast Cancer Atlas. We examined the expression of *COL12A1* in CAFs identified in human breast cancers from the Wu et al. 2021 study⁶⁷ (GSE176078). In brief, this study analysed 26 human primary breast cancers across the three major clinical subtypes (ER+, TNBC and HER2+) using single-cell RNA Sequencing on the Chromium platform (10× Genomics). Cell clusters were annotated using published gene signatures⁶³ and canonical markers (for CAFs these were *PDGFRA*, *PDGFRB*, *COL1A1*, *PDPN* and *FAP*). The expression of *COL12A1* represents log-normalised gene expression values.

Stable knockdown and overexpression of collagen XII in CAFs

Stable knockdown of collagen XII was achieved through lentiviral infection of short-hairpin constructs. Briefly, HEK293T cells were grown to 80% confluence in a 100 mm dish. Cells were transfected with 8 µg of shColXII or shSCR control lentiviral GFP constructs (OriGene, TL500400), together with 4.5 µg pMDLg/pRRE, 6.4 µg pRSV-Rev and 2.7 µg pMD.G third-generation packaging plasmids. Transfection was carried out in Opti-MEM reduced serum media using Lipofectamine 2000 as per the manufacturer's instructions. Following overnight incubation, media was replaced with DMEM supplemented with 10% FBS and 1% penicillin/streptomycin. After 24 h incubation, the lentivirus-containing media was filtered with 0.45 µm filter ready for infection. CAFs were grown to 40% confluence in DMEM supplemented with 10% FBS, 1% penicillin/streptomycin and 1% insulin-transferrin-selenium. CAFs were infected with viral supernatant in the presence of 4 µg/mL polybrene. shColXII#1, shColXII#2, shColXII#3, shColXII#4, and shSCR expressing CAFs were purified using the FACs Aria III Cell Sorter and FACs Diva software (v8.0.1, BD Biosciences,

Table 1 | gRNAs used to generate Col12a1 overexpressing cell lines

Cell Line	gRNA target sequence
eGFP-VPR	GACCAGGATGGGCACCACCC
ColXII-VPR ₁	CGGTGAAGGAGGTCTACTTT
ColXII-VPR ₂	GTAGACCTCCTCACCGCCG

USA) based on GFP expression. Briefly, CAFs were resuspended in single cell solution and washed twice with FACS buffer (PBS, 2% FBS). Following sorting, CAFs were then plated and expanded. The expression of sh constructs was routinely monitored by profiling GFP expression.

For the generation of overexpression of collagen XII, CAFs were first transduced with the dCas9-VPR-BFP construct using the lentiviral packaging system described above. BFP-positive cells stably expressing the dCas9-VPR construct were captured by FACS as described above and expanded. Lentiviral vectors carrying sgRNAs were generated by restriction cloning of gRNAs into the pLenti-sgRNA-mCherry construct using gRNAs shown in Table 1. BFP-positive cells were then transduced with the gRNA-lentiviral vector using the same packaging system described above. CAFs that were double positive for BFP and mCherry were then isolated by FAC sorting and expanded.

Collagen XII expression in stable expressing populations was confirmed by qRT-PCR and immunoblotting. Cells were routinely tested and confirmed negative for mycoplasma.

Real-time quantitative polymerase chain reaction

RNA was isolated using the Macherey–Nagel Nucleospin RNA plus kit as per the manufacturer's instruction. Reverse transcription of RNA was performed using the Quantitect Reverse Transcription Kit (Qiagen) and cDNA was synthesised from 1 µg of total RNA which was diluted 1:10. Experiments were performed using the Roche Universal Probe Library System on the Roche LightCycler 480[®] (Roche Life Sciences). *Col12a1* expression was detected with the primers forward CCAGGTCTCCTGGATATTG and reverse AAATTTGTTAGCCGGAACCTG and UPL ProbeLibrary probe 89 (#14689143001, Roche). Expression of the housekeeping gene *RPL19* was detected with primers forward CTCGTTGCCGAAAAACA and reverse TCATCCAGT-CACCTTCTCA and the UPL ProbeLibrary Probe 103 (#04692217001, Roche). *Col12a1* mRNA expression was normalised to *RPL19* expression and quantified by comparative CT as described previously⁹⁴.

Cell-derived matrices (CDMs)

CDMs were adapted from previously published protocols⁶⁹. Briefly, CAFs were plated at 2.0×10^4 cells/well in a 24-well tissue culture plate and were allowed to grow for 24 h, before media was supplemented with 50 µg/mL ascorbic acid, changed every 48 h, for 6 days.

Organotypic assays

Organotypic assays were performed as previously published⁷³ following the protocol of ref. 10 and are briefly described below.

Contraction assay. Rat-tail collagen I was extracted with 0.5 M acetic acid to a final concentration of 3.0 mg/mL. In all, 3×10^5 CAFs (shSCR, shColXII#1 or shColXII#3) were embedded in rat tail collagen I at a final concentration of 2.0 mg/mL. After polymerisation, the collagen I organotypic plugs were incubated for 12 days in DMEM supplemented with 10% FBS, 1% penicillin/streptomycin and 1% insulin–transferrin–selenium, renewing media on day 6. CAF-remodelled organotypic matrices were then either subjected to biomechanical testing or used in cancer cell invasion assays. Prior to use in invasion assays, pharmacological removal of CAFs was achieved with 400 µg/mL hygromycin for 48 h followed by

3×30 min washes in 37 °C phosphate-buffered saline, followed by 1×30 min equilibration in DMEM with 10% FBS and 1% penicillin/streptomycin to generate a cell-free matrix that had been remodelled by CAFs.

Invasion assay. Following remodelling, 1×10^5 cancer cells were seeded on top of the organotypic matrix in DMEM supplemented with 10% FBS, 1% penicillin/streptomycin, 5 µg/mL insulin, 10 ng/mL epidermal growth factor and 10 ng/mL Cholera Toxin A, and were allowed to settle for 48 h. The organotypic matrix was then transferred to a metal grid establishing an air–liquid interface and cancer cells were allowed to invade for 12 days, with the renewal of DMEM every 72 h. Organotypic matrices were then fixed in 10% formalin and processed for histological analyses. The invasive index was measured in three representative regions per organotypic matrix using the formula below (Eq. 1). Cancer cells were considered to have invaded if they were present at a distance of >50 µm from the upper surface.

$$\text{Invasive index} = \frac{\text{number of invading cells}}{\text{number of invading cells} + \text{number of non-invading cells}} \quad (1)$$

Orthotopic syngeneic in vivo study

shScr, shColXII#1 or shColXII#3 CAFs and cancer cells were resuspended as single cells and counted on an automated cell counter to obtain 1×10^6 cancer cells and 3×10^6 CAFs (1:3 ratio). Cells were mixed in 50 µL PBS and were kept on ice before being injected into the fourth mammary fatpad of 10-week-old female FVB/n mice. Tumour growth was monitored three times weekly using callipers and tumour volume was calculated using the formula:

$$\text{Tumour volume} = (\text{Maximum dimension}^2 \times \text{Minimum dimension}) \times 0.52 \quad (2)$$

The maximum tumour volume permitted by our ethics committee was 520 mm³ (Eq. 2). The maximal tumour volume was not exceeded. Secondary endpoints were defined as the development of general signs of disease/discomfort (dehydration, prolonged hunching, ruffled coat, fluid built up in the abdomen, abdominal distension, reduced movement/reactivity, obvious lesions or huddling in cage corner), or weight loss $\geq 10\%$ body weight. In this study, all mice were culled for the primary endpoint. Fresh tumour samples were subjected to biomechanical testing immediately post-collection, before being fixed in 10% formalin and processed for histological analyses. Lungs were inflated with and perfused with Fekete's solution (580 mL absolute EtOH, 80 mL 37% formalin, 40 mL glacial acetic acid and 300 mL ddH₂O) for 24 h before being processed and embedded for histological analyses. Quantification of pulmonary metastasis was performed on three-step serial sections taken 250 µm apart for each biological replicate. Metastases were defined as foci of tumour cells present in the lung above 100 µm². To quantify the size of metastases, the area of each focus was measured and normalised to the total cross-sectional area of the lung.

Statistics and reproducibility

For animal studies, group sizes were determined based on an effect size of 80%, power of 80% and alpha of 5%. An additional mouse in each group was included in the overexpression study to account for up to 10% of orthotopically injected mice that may fail to develop tumours in this model in our experience. Mice that failed to develop tumours were excluded from all analyses. Mice were randomised to treatment groups and were co-housed to minimise selection bias. For in vitro studies, the experimental design precluded any randomisation. Investigators were blinded to treatment groups for the animal experiments for tumour growth measurements. For the imaging

analysis of the orthotopic tumours and in vitro organotypic 2D and 3D matrices, all data were analysed identically using automated scripts to minimise selection bias. For all other in vitro studies, investigators were not blinded to allocation during experiments and outcome assessment due to the nature of the experimental design. The number of times experiments were independently repeated is indicated in the figure legends. All attempts to replicate the data were successful.

Unless otherwise stated, non-parametric one-way ANOVA was performed with Dunnett's multiple comparisons test, and non-parametric two-way ANOVA was performed with Tukey's multiple comparisons test. Correlations were quantified using Pearson's correlation test. Kaplan–Meier curves were analysed using a Cox Proportional Hazards Modelling for human data and a log-rank Mantel–Cox test for animal studies. Perseus (v1.6.7.0) was used for multiple sample testing of proteomic data with a false-discovery rate of 0.05. GraphPad Prism v9 was used for all other analyses. Summary data in figures are presented as mean with standard deviation unless otherwise stated. Asterisks denote statistical significance, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$ with precise p values indicated in the figure legends.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The TCGA RNAseq used in this study is publicly available from the pan-cancer dataset⁶⁰, with additional clinicodemographic information obtained from GDAC firehose and survival information from ref. 98. The publicly available human scRNA-seq data from ref. 67 is publicly available as processed scRNA-seq data for in-browser exploration and download through the Broad Institute Single Cell portal at https://singlecell.broadinstitute.org/single_cell/study/SCP1039, or through the Gene Expression Omnibus under accession number GSE176078. The publicly available mouse scRNA-seq data used in this study are available in the Gene Expression Omnibus database under accession code GSE158677⁶⁶ and https://gallegovaldeslab.shinyapps.io/pymt_shiny/. The matrisome proteomic data generated in this study have been deposited in the MassIVE <https://massive.ucsd.edu> and ProteomeXchange (www.proteomexchange.org/) databases under accession code PXD032876. The data are publicly available. The processed proteomic data are available in Supplementary Data 1. The remaining data are available within the Article, Supplementary Information or Source data file. Source data are provided with this paper.

References

- Cox, T. R. The matrix in cancer. *Nat. Rev. Cancer* **21**, 217–238 (2021).
- Kai, F., Drain, A. P. & Weaver, V. M. The extracellular matrix modulates the metastatic journey. *Dev. Cell* **49**, 332–346 (2019).
- Tian, C. et al. Cancer cell-derived matrisome proteins promote metastasis in pancreatic ductal adenocarcinoma. *Cancer Res.* **80**, 1461–1474 (2020).
- Hebert, J. D. et al. Proteomic profiling of the ECM of xenograft breast cancer metastases in different organs reveals distinct metastatic niches. *Cancer Res.* **80**, 1475–1485 (2020).
- Cox, T. R. & Erler, J. T. Remodeling and homeostasis of the extracellular matrix: implications for fibrotic diseases and cancer. *Dis. Model. Mech.* **4**, 165–178 (2011).
- Cox, T. R. & Erler, J. T. Fibrosis and cancer: partners in crime or opposing forces? *Trends Cancer* **2**, 279–282 (2016).
- Pereira, B. A. et al. CAF subpopulations: a new reservoir of stromal targets in pancreatic cancer. *Trends Cancer* **5**, 724–741 (2019).
- Maller, O. et al. Tumour-associated macrophages drive stromal cell-dependent collagen crosslinking and stiffening to promote breast cancer aggression. *Nat. Mater.* **20**, 548–559 (2021).
- Levental, K. R. et al. Matrix crosslinking forces tumor progression by enhancing integrin signaling. *Cell* **139**, 891–906 (2009).
- Calvo, F. et al. Mechanotransduction and YAP-dependent matrix remodelling is required for the generation and maintenance of cancer-associated fibroblasts. *Nat. Cell Biol.* **15**, 637–646 (2013).
- Peng, D. H. et al. Collagen promotes anti-PD-1/PD-L1 resistance in cancer through LAIR1-dependent CD8+ T cell exhaustion. *Nat. Commun.* **11**, 4520 (2020).
- Conklin, M. W. et al. Aligned collagen is a prognostic signature for survival in human breast carcinoma. *Am. J. Pathol.* **178**, 1221–1232 (2011).
- Provenzano, P. P. et al. Collagen reorganization at the tumor-stromal interface facilitates local invasion. *BMC Med.* **4**, 38 (2006).
- Bhattacharjee, S. et al. Tumor restriction by type I collagen opposes tumor-promoting effects of cancer-associated fibroblasts. *J. Clin. Invest.* **131**, e146987 (2021).
- Chen, Y. et al. Type I collagen deletion in α SMA+ myofibroblasts augments immune suppression and accelerates progression of pancreatic cancer. *Cancer Cell* **39**, 548–565.e6 (2021).
- McConnell, J. C. et al. Increased peri-ductal collagen micro-organization may contribute to raised mammographic density. *Breast Cancer Res.* **18**, 5 (2016).
- Huo, C. W. et al. High mammographic density is associated with an increase in stromal collagen and immune cells within the mammary epithelium. *Breast Cancer Res.* **17**, 79 (2015).
- Provenzano, P. P. et al. Collagen density promotes mammary tumor initiation and progression. *BMC Med.* **6**, 11 (2008).
- Martin, L. J. & Boyd, N. F. Mammographic density. Potential mechanisms of breast cancer risk associated with mammographic density: hypotheses based on epidemiological evidence. *Breast Cancer Res.* **10**, 201 (2008).
- Northey, J. J. et al. Stiff stroma increases breast cancer risk by inducing the oncogene ZNF217. *J. Clin. Investig.* **130**, 5721–5737 (2020).
- Boyd, N. F. et al. Heritability of mammographic density, a risk factor for breast cancer. *N. Engl. J. Med.* **347**, 886–894 (2002).
- Boyd, N. F. et al. Mammographic density and the risk and detection of breast cancer. *N. Engl. J. Med.* **356**, 227–236 (2007).
- Sherratt, M. J., McConnell, J. C. & Streuli, C. H. Raised mammographic density: causative mechanisms and biological consequences. *Breast Cancer Res.* **18**, 45 (2016).
- Conklin, M. W. et al. Collagen alignment as a predictor of recurrence after ductal carcinoma in situ. *Cancer Epidemiol. Biomark. Prev.* **27**, 138–145 (2018).
- Danielson, K. G. et al. Targeted disruption of decorin leads to abnormal collagen fibril morphology and skin fragility. *J. Cell Biol.* **136**, 729–743 (1997).
- Minamitani, T. et al. Modulation of collagen fibrillogenesis by tenascin-X and type VI collagen. *Exp. Cell Res.* **298**, 305–315 (2004).
- Sottile, J. & Hocking, D. C. Fibronectin polymerization regulates the composition and stability of extracellular matrix fibrils and cell-matrix adhesions. *Mol. Biol. Cell* **13**, 3546–3559 (2002).
- Mouw, J. K., Ou, G. & Weaver, V. M. Extracellular matrix assembly: a multiscale deconstruction. *Nat. Rev. Mol. Cell Biol.* **15**, 771–785 (2014).
- Shaw, L. M. & Olsen, B. R. FACIT collagens: diverse molecular bridges in extracellular matrices. *Trends Biochem. Sci.* **16**, 191–194 (1991).
- Chiquet, M., Birk, D. E., Bönemann, C. G. & Koch, M. Collagen XII: protecting bone and muscle integrity by organizing collagen fibrils. *Int. J. Biochem. Cell Biol.* **53**, 51–54 (2014).

31. Mayorca-Guiliani, A. E. et al. ISDoT: in situ decellularization of tissues for high-resolution imaging and proteomic analysis of native extracellular matrix. *Nat. Med.* **23**, 890–898 (2017).
32. Ragelle, H. et al. Comprehensive proteomic characterization of stem cell-derived extracellular matrices. *Biomaterials* **128**, 147–159 (2017).
33. Gocheva, V. et al. Quantitative proteomics identify Tenascin-C as a promoter of lung cancer progression and contributor to a signature prognostic of patient survival. *Proc. Natl Acad. Sci. USA* **114**, E5625–E5634 (2017).
34. Naba, A. et al. Characterization of the extracellular matrix of normal and diseased tissues using proteomics. *J. Proteome Res.* **16**, 3083–3091 (2017).
35. Naba, A. et al. The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Mol. Cell. Proteomics* **11**, M111.014647 (2012).
36. Rhim, A. D. et al. Stromal elements act to restrain, rather than support, pancreatic ductal adenocarcinoma. *Cancer Cell* **25**, 735–747 (2014).
37. Özdemir, B. C. et al. Depletion of carcinoma-associated fibroblasts and fibrosis induces immunosuppression and accelerates pancreas cancer with reduced survival. *Cancer Cell* **25**, 719–734 (2014).
38. Becker, L. M. et al. Epigenetic reprogramming of cancer-associated fibroblasts deregulates glucose metabolism and facilitates progression of breast cancer. *Cell Rep.* **31**, 107701 (2020).
39. Joyce, M. H. et al. Phenotypic basis for matrix stiffness-dependent chemoresistance of breast cancer cells to doxorubicin. *Front. Oncol.* **8**, 337 (2018).
40. Drain, A. P. et al. Matrix compliance permits NF- κ B activation to drive therapy resistance in breast cancer. *J. Exp. Med.* **218**, e20191360 (2021).
41. Murphy, K. J. et al. Intravital imaging technology guides FAK-mediated priming in pancreatic cancer precision medicine according to Merlin status. *Sci. Adv.* **7**, eabh0363 (2021).
42. Tian, C. et al. Suppression of pancreatic ductal adenocarcinoma growth and metastasis by fibrillar collagens produced selectively by tumor cells. *Nat. Commun.* **12**, 2328 (2021).
43. Izu, Y. et al. Collagen XII mediated cellular and extracellular mechanisms regulate establishment of tendon structure and function. *Matrix Biol.* **95**, 52–67 (2021).
44. Young, B. B., Zhang, G., Koch, M. & Birk, D. E. The roles of types XII and XIV collagen in fibrillogenesis and matrix assembly in the developing cornea. *J. Cell Biochem.* **87**, 208–220 (2002).
45. Hicks, D. et al. Mutations in the collagen XII gene define a new form of extracellular matrix-related myopathy. *Hum. Mol. Genet.* **23**, 2353–2363 (2014).
46. Veit, G. et al. Collagen XII interacts with avian tenascin-X through its NC3 domain. *J. Biol. Chem.* **281**, 27461–27470 (2006).
47. Karagiannis, G. S. et al. Proteomic signatures of the desmoplastic invasion front reveal collagen type XII as a marker of myofibroblastic differentiation during colorectal cancer metastasis. *Oncotarget* **3**, 267–285 (2012).
48. Wu, Y. & Xu, Y. Integrated bioinformatics analysis of expression and gene regulation network of COL12A1 in colorectal cancer. *Cancer Med.* **9**, 4743–4755 (2020).
49. Attalla, S., Taifour, T., Bui, T. & Muller, W. Insights from transgenic mouse models of PyMT-induced breast cancer: recapitulating human breast cancer progression in vivo. *Oncogene* **40**, 475–491 (2021).
50. Sharon, Y. et al. Tumor-derived osteopontin reprograms normal mammary fibroblasts to promote inflammation and tumor growth in breast cancer. *Cancer Res.* **75**, 963–973 (2015).
51. Goldoni, S. et al. An antimetastatic role for decorin in breast cancer. *Am. J. Pathol.* **173**, 844–855 (2008).
52. Reed, C. C. et al. Decorin prevents metastatic spreading of breast cancer. *Oncogene* **24**, 1104–1110 (2005).
53. Shen, M. et al. Tinagl1 suppresses triple-negative breast cancer progression and metastasis by simultaneously inhibiting integrin/FAK and EGFR signaling. *Cancer Cell* **35**, 64–80.e7 (2019).
54. Liot, S. et al. Loss of Tenascin-X expression during tumor progression: a new pan-cancer marker. *Matrix Biol.* **6–7**, 100021 (2020).
55. Jiang, X. et al. COL12A1, a novel potential prognostic factor and therapeutic target in gastric cancer. *Mol. Med. Rep.* **20**, 3103–3112 (2019).
56. Lattouf, R. et al. Picrosirius red staining: a useful tool to appraise collagen networks in normal and pathological tissues. *J. Histochem. Cytochem.* **62**, 751–758 (2014).
57. Junqueira, L. C., Bignolas, G. & Brentani, R. R. Picrosirius staining plus polarization microscopy, a specific method for collagen detection in tissue sections. *Histochem. J.* **11**, 447–455 (1979).
58. Chen, X., Nadiarynkh, O., Plotnikov, S. & Campagnola, P. J. Second harmonic generation microscopy for quantitative analysis of collagen fibrillar structure. *Nat. Protoc.* **7**, 654–669 (2012).
59. Williams, R. M., Zipfel, W. R. & Webb, W. W. Interpreting second-harmonic generation images of collagen I fibrils. *Biophys. J.* **88**, 1377–1386 (2005).
60. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
61. Beckers, R. K. et al. Programmed death ligand 1 expression in triple-negative breast cancer is associated with tumour-infiltrating lymphocytes and improved outcome. *Histopathology* **69**, 25–34 (2016).
62. Cazet, A. S. et al. Targeting stromal remodeling and cancer stem cell plasticity overcomes chemoresistance in triple negative breast cancer. *Nat. Commun.* **9**, 2897 (2018).
63. Wu, S. Z. et al. Stromal cell diversity associated with immune evasion in human triple-negative breast cancer. *EMBO J.* **39**, e104063 (2020).
64. Sahai, E. et al. A framework for advancing our understanding of cancer-associated fibroblasts. *Nat. Rev. Cancer* **20**, 174–186 (2020).
65. Park, D., Sahai, E. & Rullan, A. SnapShot: cancer-associated fibroblasts. *Cell* **181**, 486.e1–486.e1 (2020).
66. Valdés-Mora, F. et al. Single-cell transcriptomics reveals involution mimicry during the specification of the basal breast cancer subtype. *Cell Rep.* **35**, 108945 (2021).
67. Wu, S. Z. et al. An integrated multi-omic cellular atlas of human breast cancers. *Nat. Genet.* **81**, 129 (2021).
68. Su, S. et al. CD10+GPR77+ cancer-associated fibroblasts promote cancer formation and chemoresistance by sustaining cancer stemness. *Cell* **172**, 841.e16–856.e16 (2018).
69. Cukierman, E., Pankov, R., Stevens, D. R. & Yamada, K. M. Taking cell-matrix adhesions to the third dimension. *Science* **294**, 1708–1712 (2001).
70. Franco-Barraza, J., Beacham, D. A., Amatangelo, M. D. & Cukierman, E. Preparation of extracellular matrices produced by cultured and primary fibroblasts. *Curr. Protoc. Cell Biol.* **71**, 10.9.1–10.9.34 (2016).
71. Franco-Barraza, J., Raghavan, K. S., Luong, T. & Cukierman, E. Engineering clinically-relevant human fibroblastic cell-derived extracellular matrices. *Methods Cell Biol.* **156**, 109–160 (2020).
72. Vennin, C. et al. CAF hierarchy driven by pancreatic cancer cell p53-status creates a pro-metastatic and chemoresistant environment via perlecan. *Nat. Commun.* **10**, 3637 (2019).

73. Vennin, C. et al. Transient tissue priming via ROCK inhibition uncouples pancreatic cancer progression, sensitivity to chemotherapy, and metastasis. *Sci. Transl. Med.* **9**, eaai8504 (2017).
74. Chitty, J. L. et al. The Mini-Organ: a rapid high-throughput 3D coculture organotypic assay for oncology screening and drug development. *Cancer Rep.* **3**, e1209 (2020).
75. Cicchi, R. et al. Scoring of collagen organization in healthy and diseased human dermis by multiphoton microscopy. *J. Biophotonics* **3**, 34–43 (2010).
76. Timpson, P. et al. Organotypic collagen I assay: a malleable platform to assess cell behaviour in a 3-dimensional context. *J. Vis. Exp.* e3089 (2011).
77. Sun, M. et al. Collagen XII regulates corneal stromal structure by modulating transforming growth factor- β activity. *Am. J. Pathol.* **192**, 308–319 (2022).
78. Schönborn, K. et al. Role of collagen XII in skin homeostasis and repair. *Matrix Biol.* **94**, 57–76 (2020).
79. Floerchinger, A. et al. Optimizing metastatic-cascade-dependent Rac1 targeting in breast cancer: Guidance using optical window intravital FRET imaging. *Cell Rep.* **36**, 109689 (2021).
80. Wang, S. et al. CCM3 is a gatekeeper in focal adhesions regulating mechanotransduction and YAP/TAZ signalling. *Nat. Cell Biol.* **23**, 758–770 (2021).
81. Alexander, J. & Cukierman, E. Stromal dynamic reciprocity in cancer: intricacies of fibroblastic-ECM interactions. *Curr. Opin. Cell Biol.* **42**, 80–93 (2016).
82. Beacham, D. A. & Cukierman, E. Stromagenesis: the changing face of fibroblastic microenvironments during tumor progression. *Semin. Cancer Biol.* **15**, 329–341 (2005).
83. Tian, C. et al. Proteomic analyses of ECM during pancreatic ductal adenocarcinoma progression reveal different contributions by tumor and stromal cells. *Proc. Natl Acad. Sci. USA* **116**, 19609–19618 (2019).
84. Yue, W. et al. Fibulin-5 suppresses lung cancer invasion by inhibiting matrix metalloproteinase-7 expression. *Cancer Res.* **69**, 6339–6346 (2009).
85. Naba, A., Clauser, K. R., Lamar, J. M., Carr, S. A. & Hynes, R. O. Extracellular matrix signatures of human mammary carcinoma identify novel metastasis promoters. *Elife* **3**, e01308 (2014).
86. Sun, M. et al. Collagen XI regulates the acquisition of collagen fibril structure, organization and functional properties in tendon. *Matrix Biol.* **94**, 77–94 (2020).
87. Wipff, P.-J., Rifkin, D. B., Meister, J.-J. & Hinz, B. Myofibroblast contraction activates latent TGF- β 1 from the extracellular matrix. *J. Cell Biol.* **179**, 1311–1323 (2007).
88. Sun, X. et al. Tumour DDR1 promotes collagen fibre alignment to instigate immune exclusion. *Nature* **599**, 673–678 (2021).
89. Nicolas-Boluda, A. et al. Tumor stiffening reversion through collagen crosslinking inhibition improves T cell migration and anti-PD-1 treatment. *Elife* **10**, e58688 (2021).
90. Guy, C. T., Cardiff, R. D. & Muller, W. J. Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease. *Mol. Cell. Biol.* **12**, 954–961 (1992).
91. Hughes, C. S. et al. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat. Protoc.* **14**, 68–85 (2019).
92. Bruderer, R. et al. Optimization of experimental parameters in data-independent mass spectrometry significantly increases depth and reproducibility of results. *Mol. Cell. Proteomics* **16**, 2296–2309 (2017).
93. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
94. Conway, J. R. W. et al. Three-dimensional organotypic matrices from alternative collagen sources as pre-clinical models for cell biology. *Sci. Rep.* **7**, 16887 (2017).
95. Zeltz, C. et al. α 11 β 1 Integrin is induced in a subset of cancer-associated fibroblasts in desmoplastic tumor stroma and mediates in vitro cell migration. *Cancers* **11**, 765 (2019).
96. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297 (2012).
97. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
98. Liu, J. et al. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* **173**, 400.e11–416.e11 (2018).
99. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).
100. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
101. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. Preprint at <https://arxiv.org/abs/1802.03426> (2020).

Acknowledgements

This project was supported by the National Health and Medical Research Council (NHMRC) project (1129766, 1140125, 2000937; T.R.C.) and fellowship (1158590; T.R.C.) funding, a Cancer Institute NSW Career Development Fellowship (CDF171105; T.R.C.), Cancer Council NSW grants (RG19-09; T.R.C.), Susan G Komen for the Cure (CCR17483294; T.R.C.), Perpetual IMPACT funding (IPAP2020/0066; T.R.C.), Love Your Sister in association with the Australian National Breast Cancer Foundation (IIRS-21-041; T.R.C.), Suttons family support (P.T.), a National Health and Medical Research Council Senior Research Fellowship (1136974; P.T.) and Australian Cancer Research Foundation (ACRF) INCiTe Centre (P.T.) funding. D.H., B.P. and K.J.M. were supported by Cancer Institute NSW Early Career Fellowships (2018/ECF011 to D.H.; 2021/ECF1309 to B.P.; 2021/ECF1384 to K.J.M.). The authors thank the various members of their labs for the critical reading of the manuscript.

Author contributions

T.R.C. and P.T. conceived, designed, and directed the project with the support of M.P., A.L.P., E.L., S.O.T. and B.G.O. M.P. and A.L.P. as co-first authors executed the project, assisted by M.Y., E.C.F., S.Z.W., J.L.C., K.W., E.T., E.M., A.N., J.N.S., M.C.L., D.H., M.N., B.A.P., A.L., L.C., K.J.M., A.Z., J.F.H., D.R.C., E.L., B.G.O., F.V.M., B.L.P., D.G.O., A.S. and S.O.T. T.R.C., M.P., M.C.L., J.N.S., E.C.F., D.H. and B.L.P. optimised the decell proteomics methodology necessary for mass spectrometry profiling. B.L.P. prepared and carried out mass spectrometry profiling of PyMT tissue samples. M.P. analysed and interrogated mass spectrometric data under the direction of T.R.C. and with support from B.L.P. M.P. carried out multiphoton (SHG) imaging and characterisation of samples. M.P., assisted by A.L.P., performed second harmonic generation imaging and quantification of cell-derived matrices. S.Z.W. and A.S. completed single-cell sequencing analysis of collagen 12 in the human breast cancer cohort that forms part of the Human Breast Cancer Atlas. F.V.M. and D.G.O. completed single-cell sequencing analysis of tumours from the PyMT model for collagen 12. M.P. with the assistance of A.L.P. generated and validated Col12a1 knockdown and overexpression lines. M.P. conducted, scored and analysed the in vitro cell-derived matrix (CDM) experiments. M.P. together with A.L.P., E.M., E.T. and A.N. conducted, scored and analysed in vitro 3D organotypic assays. M.P. and T.R.C. undertook biomechanical profiling and analysis of in vitro and in vivo

samples. M.P. and T.R.C. carried out in vivo orthotopic co-implantation animal models supported by M.Y., E.T., E.M., A.N. and K.W. M.P. undertook histological assessment of mouse tissue samples. Histologically stained and Picrosirius Red-stained samples were analysed using in-house scripts written by T.R.C. and interpreted by M.P., A.L.P., T.R.C. and P.T. Scoring and analysis of pulmonary metastatic burden in animal models was carried out by M.P. M.P., M.Y., A.L.P. and K.W. completed western blotting studies presented in the manuscript. M.P. optimised human TMA staining protocols. S.O.T. scored human TMAs and assisted A.L.P. in the analysis of clinicopathological follow-up data. A.L.P., assisted by M.P., completed the univariate survival analysis of TCGA data. A.L.P. completed the analysis of COL12A1 expression in tumour and non-tumour tissue from the TCGA cohort. A.L.P. as assisted by M.P. performed multivariate analyses of patient survival in the TCGA and TMA cohorts, including interpretation of clinical co-variables, with guidance from S.O.T. A.L.P. assessed the relative contributions of COL12A1 and CAF signatures to patient survival. M.P., under guidance from T.R.C. and P.T., prepared the initial manuscript for submission. A.L.P. and M.P., under guidance from T.R.C. and P.T., revised the manuscript and responded to reviewers following peer review.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at

<https://doi.org/10.1038/s41467-022-32255-7>.

Correspondence and requests for materials should be addressed to Paul Timpson or Thomas R. Cox.

Peer review information *Nature Communications* thanks Jasper Wouters, Sara Zanivan and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022